THE ROLE OF THE LOMBARD REFLEX IN THE PRODUCTION AND PERCEPTION OF CANTONESE-ACCENTED ENGLISH

by

Chi-nin Li

B.Sc., University of Victoria, 1996

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Linguistics

SIMON FRASER UNIVERSITY

November 2000

0-612-61453-0

Abstract

The Lombard reflex is a phenomenon that occurs when speakers modify their vocal effort while speaking under noisy conditions. When presented at an equal level of masking noise, speech originally produced in noise (i.e., Lombard speech) is more intelligible than speech produced in quiet (i.e., normal speech). Because research associated with the Lombard effect has been conducted mainly with native speakers, little is known about its impact on foreign-accented speech. This study examines to what extent the Lombard reflex affects the production and perception of Cantonese-accented English.

Twelve native Hong Kong Cantonese speakers and a comparison group of Canadian English speakers participated in the production task. The speakers read a set of simple true and false English statements in quiet and in 70 dB of cafeteria-like masking noise. Some productions of the speakers were subjected to acoustical measurements and selected for a subsequent perceptual task. The acoustic analyses revealed that for both groups of speakers, sentential fundamental frequency was significantly different in speech produced in quiet and in noise. However, the

native English speakers reduced their speaking rates to a greater extent in noise than did the Cantonese speakers.

Selected normal and Lombard utterances were duplicated and mixed with noise at a constant signal-to-noise ratio, and presented along with noise-free stimuli to a new group of eight native English listeners who evaluated the utterances along three dimensions. They assessed intelligibility by transcribing the sentences and verifying their truth value, and also assigned foreign-accent and comprehensibility ratings on 9-point scales. The Cantonese speakers' utterances were found to be more accented, but less intelligible and less comprehensible than those of the English speakers in all conditions. In addition, the Cantonese speakers' Lombard sentences were correctly transcribed more often than their normal utterances in noisy conditions. However, their Lombard speech was as intelligible and comprehensible as normal utterances when presented without masking noise. Accent ratings, interestingly, were similar throughout all conditions.

The findings of this research suggest that manifestations of the Lombard reflex on first language production are different from those on second language production, and that the Lombard reflex is a highly speaker-dependent phenomenon that is affected by the native language of the speakers.

Acknowledgements

I would like to express my deepest gratitude to my senior supervisor, Dr. Murray Munro. He was always there to help me deal with all the problems that I encountered. I am grateful for his reading and commenting on all the drafts of this thesis. Without his continuous encouragement and direction, this research would not have been possible. In addition, this research was supported through a grant by the SSHRC to Dr. Munro.

I would also like to thank the other members of my examining committee, Dr. Zita McRobbie and Dr. Susan Guion, for their insightful and valuable comments. Special thanks go to Dr. Wyn Roberts, Dr. Nancy Hedberg, and Dr. Megan Hodge (U. of Alberta) for their advice on a number of issues in this research. I thank Mr. Dave Murphy (School of Communication), Mr. Jeff Rudd (Department of Physics), and Mr. Richard Chi Fai Li (Microcomputer Store) for their technical assistance and suggestions.

I would like to extend my thanks to all my fellow graduate students for their support and friendship, especially those who participated in my experiment, and to the always friendly staff in Linguistics. I offer thanks to Dr. Roberts, Mr. Cliff Burgess, and Miss Kyoung-Ja Lee for their constant encouragement. Special thanks go to all participants for their contribution to my thesis.

I am also very grateful to Connie, as she has been an unfailing source of support, help, and encouragement throughout these years.

Last but not least, thank the Lord for guiding me through all the difficulties and frustrations.

Table of Contents

List of Tables

List of Figures

CHAPTER I

INTRODUCTION

In general, people speak much louder in noisy conditions such as a busy cafeteria than in a quiet environment such as a library. This phenomenon and its acoustic-phonetic consequences are known as the Lombard reflex. Its characteristics have been evaluated and reported in a number of studies, primarily of speech produced by native English speakers. Speech originally produced in a background of noise has been reported to be more intelligible than speech originally produced in quiet conditions, when both types of speech are mixed with the same level of masking noise and presented to listeners. Findings from this kind of research are important to the analyses of speech communication and speech intelligibility in real-world situations. There have been anecdotal claims that loud ambient noise may interfere with one's understanding of the conversation, especially when foreign-accented speech is involved. However, there has been very little published literature on this issue.

Almost all the studies associated with speech communication in noise have been carried out with native speakers, whereas studies of second language speech have mainly been carried out under ideal laboratory conditions that hardly represent realistic communicative situations. To fill the gap in this area of research, the present study examines the production and perception of Cantonese-accented English under adverse listening conditions.

The purpose of this chapter is to discuss the general properties of the Lombard reflex and its effects on speech perception. A review of the proposed underlying mechanisms accounting for the phenomenon is followed by a section on purposes for this thesis.

1.1. THE LOMBARD REFLEX

The Lombard reflex is a phenomenon according to which, normal-hearing speakers will unconsciously increase their vocal levels in the presence of a loud background noise (Anglade & Junqua, 1990; Applebaum, Hanson, & Morin, 1996; Bond, Moore, & Gable, 1989; Dreher & O'Neill, 1957; Egan, 1971; Junqua, 1993, 1996; Junqua & Anglade, 1990; Lane & Tranel, 1971; Van Summers, Pisoni, Bernacki, Pedlow & Stokes, 1988). This phenomenon was first described in 1911 by Etienne Lombard, the first oto-rhino-laryngologist in France (Lane & Tranel, 1971).

According to Egan (1971), Lombard suddenly presented an intense noise bilaterally to normal hearing patients who were reading aloud, or answering simple questions, and observed large increases in patients' vocal levels. Lombard suggested that, because of the interfering effects of the masking noise, normal hearing individuals behave briefly like people with sensorineural hearing loss (Egan, 1971). However, when the noise stops, the voice intensity returns back to a normal level (Egan, 1971; Junqua, 1996; Lane & Tranel, 1971).

1.2. CHARACTERISTICS

Lane and Tranel (1971) and more recently, Junqua (1996), have conducted detailed surveys of the literature on the Lombard reflex phenomenon and its associated characteristics manifested in speech. Lombard speech has been reported to be different from normal speech in a number of ways. For instance, increases in voice level, fundamental frequency and vowel duration, and a shift in formant center frequencies for F1 and F2 have been documented (Anglade & Junqua, 1990; Applebaum et al., 1996; Junqua, 1996; Junqua & Anglade, 1990). In addition to the above characteristics, it has once been reported that speaking rate may be reduced when speech is produced in a noisy environment (Hanley & Steer, 1949).

However, the evidence for this feature of the Lombard reflex has been controversial (Lane & Tranel, 1971). Bond et al. (1989) concluded that all the changes associated with Lombard speech reflect articulatory modifications made to increase the vocal effort and to articulate in a more precise manner for better communication in a noisy condition.

It has been noted that very little is known about the acoustic changes that occur in speech in a noisy environment, because the manifestations of this phenomenon are different from person to person (Junqua, 1996; Van Summers et. al., 1988). For instance, Van Summers et al. (1988) reported that a significant increase in fundamental frequency was observed for one, but not both, of their male speakers, when they spoke in quiet and in different levels of noise. Junqua (1996) agreed that the nature of the Lombard reflex is highly speaker-dependent. Moreover, the manifestation of the Lombard reflex may also vary with the type of ambient noise, and with the language of the speaker (Junqua, 1996). Lane and Tranel (1971) have suggested that the magnitude of the speakers' response to noise seems to be governed by the desire to achieve intelligible communication. They argued, for example, that in a noisy condition, speakers would not change their voice level when talking to themselves. In agreement with Lane and Tranel (1971), Bond et al.

(1989) observed that the magnitude of the Lombard effect is greater when speakers believe they are communicating with interlocutors. Given this complex set of factors, the Lombard reflex is not thought of as an all-or-none response with a certain threshold and relatively fixed magnitude (Junqua, 1996; Lane & Tranel, 1971). In other words, the binary situation, Lombard versus non-Lombard speech, does not exist (Junqua, 1996). As pointed out by Junqua (1996), the variability in Lombard speech appears to be distributed along a continuum.

Acoustic differences between Lombard speech and normal speech seem to have an effect on intelligibility. In perceptual experiments, for instance, speech produced in noise is more intelligible, and thus better perceived, than speech produced in quiet, when both types of speech are presented in noise at an equivalent signal-to-noise ratio (Dreher & O'Neill, 1957; Junqua, 1993; Van Summers et al., 1988). However, if the utterance reaches a level of shouting, around 110 to 120 dB, the intelligibility decreases at any constant S/N ratio (Pollack & Pickett, 1958). In addition, Junqua (1996) noted that the type of masking noise and the gender of the speakers used for the experiment are crucial to the difference in intelligibility of speech produced in noise-free and in noisy conditions. Multitalker noise is found to degrade the intelligibility of English digit vocabulary more than white noise (Junqua, 1993). Moreover, Junqua (1993) reported that when presented in multitalker noise, female Lombard speech is more intelligible than is male Lombard speech. As pointed out by Junqua (1993), breathiness decreases the intelligibility of speech. It seems that when producing speech in a noisy background, female speakers tend to decrease the breathiness in their productions more than male speakers do (Junqua, 1993).

1.3. MECHANISMS

Researchers who have tried to account for the Lombard reflex hold a variety of different perspectives. On the one hand, it has been claimed that the phenomenon is due to an automatic auditory regulating device (Fairbanks, 1954; Fricke, 1970; Lombard, 1911: cited in Junqua, 1993; Rivers & Rastatter, 1985). On the other, there is an emphasis on the speakers' response to the listeners' needs (Fonagy & Fonagy, 1966; Lane & Tranel, 1971). For some researchers (Junqua, 1996; Van Summers et al., 1988), the mechanism for the Lombard reflex is viewed as a combination of the above two perspectives.

Auditory feedback has been considered to be an element of the human speech production system (Rivers & Rastatter, 1985).

Fricke (1970) suggest that the mechanism responsible for the Lombard reflex is a result of noise being introduced into the auditory feedback channel. This idea is based on a model proposed by Fairbanks (1954) in which the human communication mechanism is viewed as a servosystem, a kind of self-regulatory feedback system. In brief, a sample of the output is fed back to some point in the speaking system where the ensuing output is controlled. According to Fairbanks' (1954) model, one's speech output (i.e., one's speech signal being masked with noise, in this case) is monitored at any point in time. This auditory feedback signal is then compared with an intended input signal. Any difference between these two signals will lead to an adjustment to the output signal. The loop of comparison continues until the differential reduces to zero. As a consequence, the ultimate output will be adjusted to be the same as the desired input signal.

However, Lane and Tranel (1971) claimed that the proposal that the human speaking mechanism is a servomechanism is probably a misinterpretation of the Lombard reflex, and overlooks the communicative function involved. Instead, they advocated that in a background of masking noise, the magnitude of a speaker's response is more or less governed by the premium on intelligible communication. In their opinion, the notion of auditory feedback control is insufficient to fully describe the phenomenon. For example, when noise is presented to one ear only, the voice level still increases even though the feedback control can be processed through the other undisturbed ear. They argued that in a noisy environment, the speaker exhibits Lombard speech in order to be understood by the listener or interlocutor during communication. An increase in the ambient noise, or a decrease in the level at which a speaker is heard, implies a reduction in the signal-to-noise ratio of the utterance relative to the environment. In adverse conditions, the speaker suffers from reduced intelligibility. Lane and Tranel (1971) asserted that the reason for exhibiting Lombard speech is to compensate for the reduced S/N ratio so as to ensure intelligible communication, and that the extent of the compensatory strategies is related to how demanding the communicative situation is. Nevertheless, as suggested by Black (1954), the magnitude of responses may also be dependent on how much the speaker feels responsible for the communication. Fonagy and Fonagy (1966) concluded that all changes in articulation (e.g., increases in voice level, vowel duration, or pitch) serve to maintain an intelligible communication in adverse communicative conditions.

To Junqua (1996) and Van Summers et al. (1988), the most plausible explanation for the Lombard reflex, however, appears to be a combination of the above two ideas.

1.4. SPEECH PRODUCTION IN NOISE

Among all the features associated with speech produced in noise, a change in voice level is the most obvious manifestation of the Lombard effect. It is widely agreed by researchers that when speakers talk in a background of noise, their speech become more intense (Bond et al., 1989; Lane & Tranel, 1971; Steeneken & Hansen, 1999). The majority of studies of the Lombard reflex in the past decades has been concerned with the relation between the voice intensity level and the masking noise level (Dreher & O'Neill, 1957; Gardner, 1964, 1966; Hanley & Steer, 1949; Korn, 1954; Kryter, 1946; Pickett, 1956; Webster & Klump, 1962). However, research examining other characteristics, such as speaking rate or fundamental frequency, of the Lombard effect seem relatively few in number.

In summarizing intelligibility tests on words produced in high-level aircraft noise for the US Air Forces in the mid 40s, Haagen (1946) concluded that particularly slow and fast speech rates are detrimental to intelligibility. However, a lack of further detailed research on speaking rate and its relation to speech intelligibility was probably due to non-significant results of some sentence tests at the time (Haagen, 1946). Unfortunately, Haagen (1946) offered no explanation for that. Later, Hanley and Steer (1949) examined the effects of four levels of aircraft noise (about 100dB, 106dB, 114dB, 122dB) on speaking rate in a passage read by 48 male American English speakers. They found that the higher the noise level was, the slower was the speaking rate. The mean speaking rate (in words per minute) dropped successively from 183.2 to 165.4. However, this line of research appears not to have been pursued.

In contrast, another characteristic of the Lombard reflex, fundamental frequency, has been investigated in a number of studies. An increase in Fo has been consistently reported when English words are read in citation form in a background of masking noise (Anglade & Junqua 1990; Bond et al., 1989; Junqua, 1993; Junqua & Anglade, 1990; Van Summers et al., 1988). Instead of using isolated words as reading materials, Rivers and Rastatter (1985) examined the fundamental frequency of stressed and unstressed English words in spontaneous speech produced in quiet and noisy conditions. For the latter conditions, two types of noise each at 90 dB were used: white noise and multitalker noise (i.e., simultaneous conversations of a crowd of people). Ten adults (5 female, 5 male) and 28 English-speaking children, the latter being divided into three age groups with balanced gender (5-year-, 7-year-, and 10-year-olds), served as speakers in that experiment. Rivers and Rastatter (1985) found that, overall, their speakers exhibited higher mean Fo values when speaking in the two masking conditions than when speaking in the quiet condition. The children speakers' mean Fo values (pooled over stressed and unstressed words) observed in white noise (273.7 Hz for females and 299.8 Hz for males) and in multitalker noise (284.2 Hz for females and 299.7 Hz for males) were much higher than in quiet (241.7 Hz for females and 254.7 Hz for males). As was the case for the children, the adult participants utilized mean Fo values that were lower in quiet (223.5 Hz for females and 109.5 Hz for males) than in either white noise (248.5 Hz for females and 122.0 Hz for males) or multitalker noise (253.0 Hz for females and 128.0 Hz for males). Loren et al. (1986) replicated the experiment by Rivers and Rastatter (1985) in measuring the effects of noise on fundamental frequency in extemporaneous speech produced by American English speakers. However, only young adults (3 female, 3 male) were used as subjects. They described a series of pictures spontaneously, once in quiet and once in a background of white noise at 90 dB. In that study, the mean Fo for each sentence, rather than the mean Fo for individual words in sentences (Rivers & Rastatter, 1985), was measured and compared for the two speaking conditions. Loren et al. (1986) found that all subjects utilized sentential Fo values higher in the noisy condition (261.9 Hz for females and 127.1 Hz for males) than in the quiet condition (220.6 Hz for females and 105.6 Hz for males).

1.5. SPEECH PERCEPTION IN NOISE

As already mentioned, speech produced in a noisy environment has been reported to be more intelligible than quiet-produced speech when presented at a constant signal-to-noise ratio. Pisoni (1996) noted that intelligibility has commonly been evaluated by presenting words masked with noise to listeners for identification. Similar studies of word identification in noise have been carried out by Dreher and O'Neill (1957), Hazan and Simpson (1998), Junqua (1993), Junqua and Anglade (1990), Lane (1963), Pollack and Pickett (1958), and Van Summers et al. (1988). The latter found that at an equivalent S/N ratio, the English digit vocabularies (from zero through nine) produced in 90 dB and 100 dB of masking noise were correctly identified more often than those produced in a quiet condition. However, it is noteworthy that Dreher and O'Neill (1957) earlier tested the intelligibility not only of words, but also of sentences, produced by speakers in quiet and various levels of masking noise. In their experiment, a group of English speakers (12 female, 3 male) read five words and five sentences in each of five speaking conditions. For each of the speakers, utterances were recorded in quiet and in a background of white noise at four levels (70, 80, 90, and 100 dB) presented to them via headphones. The recorded speech was equated for overall intensity and then mixed with wideband random noise at a constant S/N ratio of +4 dB. The processed stimuli were presented and identified by 204 English-speaking listeners in an orthographic transcription task. It was found that, first of all, the unmasked normal words and sentences were highly intelligible (98% on average). In addition, the Lombard utterances were found to be more intelligible than normal speech when presented in the masking noise. Also, Dreher and O'Neill (1957) reported that sentences were always more intelligible than isolated words, and suggested that the difference in intelligibility was probably attributable to the sentence context. It had already been noted by Miller, Heise, and Lichten (1951) that word intelligibility is higher in a sentence context than in citation form.

It appears that studies of the perception of Lombard or normal speech presented in noise have been mostly concerned with L1 speakers, especially native English speakers. However, the production and perception of foreign-accented speech in noise have been given relatively little attention. Lane (1963) can be considered one of the pioneers in studying foreign-accented speech perceived under adverse listening conditions. He examined the effects and interactions of two types of speech distortion: response-independent and response-dependent. According to Lane (1963), response-independent distortion is defined as any speech distorting operation (e.g., masking with noise) applied to a speech signal that is not determined by the response of the listener. In contrast, response-dependent distortion (e.g., foreign accent) relies on the listener's response during undistorted speech transmission. Moreover, the distinction between the two types of operations can also be viewed as whether the distortion comes from sources external (e.g., filtering in speech communication channels) or internal to the speakers (e.g., aphasic speech).

In Lane's (1963) experiment, lists of English monosyllabic words produced by 1 native English speaker and by 3 nonnative speakers with a strong accent in English (Serbian, Punjabi, and Japanese) were recorded. To create response-independent distortion, white noise was mixed with the recorded speech signals at four S/N ratios, +15, +4, -1.5, and –5dB. In order to assess the intelligibility of the utterances, the processed signals were presented to a group of native English-speaking listeners. Lane (1963) found that the correct identification scores for all speakers' productions decreased as the level for the masking noise increased. Moreover, at all four S/N ratios, the nonnative utterances were much less intelligible (about 36%, overall) than were those of the native speech, with no interaction effects between response-independent and response-dependent operations. The latter finding was important in that the effect of foreign accent on speech intelligibility was found to be independent of the effect of masking. Lane (1963) also found minimal individual differences among the nonnative speakers of English used in his experiment.

In a more recent study, Munro (1998) examined the effects of cafeteria noise on the perception of English produced by groups of native speakers of Mandarin and of English. In brief, speakers read a list of 40 simple true and false English sentences in an ideal laboratory recording condition. Some of the recorded utterances from each speaker were duplicated and mixed with cafeteria noise at a mean S/N ratio of +7.9 dB. The mean root-mean-square (RMS) amplitudes for all unmasked and masked sentences were equated and presented to a new group of 24 native English-speaking listeners. The intelligibility of the sentences was assessed through a sentence-verification task and a sentence-transcription task. The findings suggested that the cafeteria noise did degrade the intelligibility of the sentences overall. Munro (1998) also found that in both tasks, Mandarin speakers' sentences were less intelligible than those of the native productions in both noise-free and noisy conditions. In addition, even for those utterances that were completely intelligible in the noise-free condition, the Mandarin accented sentences were more susceptible to the masking noise (i.e., less intelligible) than were the statements produced by the native English speakers. An analysis of mistranscription errors made by the listeners revealed that the errors in the noise-free Mandarin speakers' utterances were much more frequent than those for the noisy native productions. As noted by Munro (1998), the qualitatively different effects of accent and noise on speech intelligibility supported the aforementioned distinction by Lane (1963) between response-dependent and response-independent distorting operations. However, in contrast to what Lane (1963) had found, large interspeaker variability was observed among the Mandarin speakers.

Nonetheless, neither of these two studies, Lane (1963) and Munro (1998), looked at the effects of noise on the production of foreign-accented speech. In other words, different results and additional information might have been obtained if they had presented Lombard speech, along with normal speech, produced by their L2 speakers to the native English listeners.

1.6. PURPOSES OF RESEARCH

The main purpose of this study is to explore the effects of noise on the production and perception of nonnative English speech. As noted by Lane and Tranel (1971), the Lombard reflex has been considered an important phenomenon in the analysis of speech communication in noise. However, almost all the previous studies associated with the Lombard reflex have been carried out with native English speakers. Moreover, because numerous studies of the speech of second language (L2) speakers have been carried out under ideal laboratory conditions, the production and perception of nonnative speech under nonideal circumstances have received relatively little attention.

In this study, an attempt is made to examine the production and perception of Lombard speech produced by Hong Kong Cantonese speakers of English. In the production study, two acoustic parameters associated with the Lombard reflex, speaking rate and fundamental frequency, are examined and compared in sentences produced under normal and noisy conditions by native Cantonese and native English speakers.

In the perception study, the sentence-length utterances are presented with or without masking noise to native English-speaking listeners for evaluations along three dimensions applicable to foreign-accented speech: intelligibility, foreign-accentedness, and comprehensibility. Intelligibility, a crucial aspect of L2 speech (Derwing & Munro, 1997), refers to the extent to which a nonnative utterance is understood by a listener (Munro & Derwing, 1995a, 1995b, 1999). The notion of accentedness is defined as the degree to which the speech of nonnative speaker is perceived to be different from native speakers' typical productions (Munro, 1998; Munro & Derwing, 1995b, 1998, 1999). The third dimension, comprehensibility, is defined as the degree of difficulty perceived by listener in understanding the nonnative speech (Derwing & Munro, 1997; Munro & Derwing, 1995b, 1998, 1999). The present study aims to investigate the extent to which the nonnative Lombard speech will be perceived to be different on these three dimensions from the corresponding normal speech presented in quiet and noisy conditions. The findings are also discussed in connection with the results obtained in the preceding production experiment.

An examination of previous studies including those summarized above indicates that research associated with the Lombard reflex has been done under well-controlled experimental conditions that can hardly represent real-world situations. In most studies (Anglade & Junqua, 1990; Applebaum et al., 1996; Bond et al., 1989; Dreher & O'Neill, 1957; Fricke, 1970; Hanley & Steer, 1949; Junqua, 1993; Junqua & Anglade, 1990; Loren et al., 1986; Van Summers et al., 1988), participants read the stimulus materials alone in the studio without any listener or interlocutor. In this research, to be more realistic within a laboratory setting, speaker-participants must listen carefully to instructions given by the experimenter throughout all recording sessions, and they are requested to produce sentences in a conversation-like manner.

CHAPTER II

EXPERIMENT 1: EFFECTS OF NOISE ON CANTONESE SPEAKERS' ENGLISH SENTENCE PRODUCTION

The purpose of this experiment is to examine the effects of the Lombard reflex on sentences spoken by Cantonese speakers of English. In a detailed review of the literature about the phenomenon by Lane and Tranel (1971) and a more recent one by Junqua (1996), it is observed that previous studies concerning the Lombard reflex over the past few decades have generally focused on speech produced by speakers of native languages such as English, French, or Spanish (Junqua, 1996). However, little is known about the effects of noise on speech produced by nonnative speakers of language.

In this study, twelve native Hong Kong Cantonese speakers (6 female, 6 male) were asked to read aloud in a conversation-like manner to the experimenter in a highly instruction-based recording session a list of 48 simple declarative English sentences, once in a quiet condition and once in a noisy condition. In the latter, cafeteria-like masking noise at 70 dB was presented to the speakers' headphones. A group of twelve native Canadian English speakers balanced for gender served as a comparison group. Speaking rate and fundamental frequency (mean, and Fo range) values were computed from four different productions from each speaker, identical in the two speaking conditions. Findings from previous studies of the Lombard effect indicate that in a noisy environment, the speaking rate for utterances will decrease (Hanley & Steer, 1949), while fundamental frequency will increase (Bond et al., 1989; Loren et al., 1986; Rastatter & Rivers, 1983; van Summers et al., 1988). It is thus expected that the subjects in this experiment will tend to exhibit slower speech rates, and utilize higher fundamental frequency values in noisy than in quiet conditions.

In addition, results from studies of second language speech show that the speech rates for nonnative speakers of English are generally slower than those for native English speakers (Anderson-Hsieh & Koehler, 1988; Guion, Flege, Liu & Yeni-Komshian, 2000; Munro, 1995; Munro & Derwing, 1998, 1999), and that Mandarin speakers of English exhibit higher fundamental frequency values than do their native counterparts (Munro, 1995). In view of the above, it is reasonable to anticipate that the Cantonese speakers, like other L2 speakers of English, will produce the sentence-length utterances at a slower rate than will the native English talkers. Also, since both Mandarin and Cantonese are Chinese languages, the Cantonese speakers, like the Mandarin speakers, are expected to exhibit higher fundamental frequency values (mean Fo and Fo range) than those of the native speakers of English. Interactions between the speaking conditions and the native language of the speakers for the two respective measures will be investigated. The results of this production experiment will be interpreted in terms of characteristics of the native languages of the speakers and of the Lombard reflex.

2.1. METHOD

2.1.1. Participants

Twelve native Hong Kong Cantonese speakers, six males and six females, were recruited as subjects. They were all born and raised in Hong Kong, and had not lived for an extended time in any English speaking area besides Canada. Before coming to Canada, they had grown up in Hong Kong, receiving education from kindergarten to at least part of Form Five (the highest grade in high school). All participants started learning English in kindergarten. They had a mean age of 21.3 years at the time of study with a range of 20 to 26 years. They had lived in Canada for a mean of 25.1 months with a range of three weeks to 45 months. They were undergraduate students either at Simon Fraser University or at community colleges in Vancouver, British Columbia. The speakers were selected to ensure that all of them had some degree of Hong Kong Cantonese accent in English and that a range of degrees of accent were represented.

Twelve native speakers of English (6 females, 6 male) served as a comparison group. All the English participants were speakers of Canadian English who had been born and raised in British Columbia. They were either graduate or undergraduate students at Simon Fraser University. The mean age was 23.3 years with a range from 19 to 30 years. None of the speakers in either group exhibited any voice or speech anomalies.

All speakers first completed a Language Background Questionnaire (LBQ) giving general information about their language background (see Appendix 1 for Cantonese speakers and Appendix 2 for English speakers). They all passed a pure-tone hearing screen (250, 500, 1000, 2000, and 4000 Hz at 25 dB) binaurally with the use of a Maico MA 25 audiometer prior to performing the production tasks. They had not previously participated in any speech experiment involving noise. None had heard of the Lombard reflex, except a female native English speaker who was a graduate student in the Department of Linguistics of Simon Fraser University. She claimed to have heard of the term, but she knew nothing about it. All speakers participated for approximately one hour, and each of them was paid a $15 honorarium.

2.1.2. Stimulus Sentences

A list of 24 true and 24 false sentences served as the basis for the study (see Appendix 3). Half of the true and half of the false statements among these 48 sentences had been used in previous experiments (Derwing, Munro, & Wiebe, 1998; Munro, 1998; Munro & Derwing, 1995b, 1999). The remainder were constructed by the experimenter. Each item was a single clause sentence of four to eight words. The number of syllables ranged from five to 12. All vocabulary items in the statements were listed as high frequency words by Sakiey and Fry (1979).

The set of 48 statements was randomly printed on three pieces of paper. To avoid any possible list-initial and list-final effects on reading (Cooper, Eady, & Mueller, 1985; Leder & Spitzer, 1993), the fourth and the fifth statements were duplicated, and arranged to become the last and the first sentence on each page. These two sentences were not used for acoustic measurements nor for subsequent perception tasks. Thus, there were ten statements altogether on each side of the stimulus sheet. All pieces of paper were laminated, and were put on a desktop document holder that was placed at a distance where each speaker could clearly see the sentences. Six different randomized lists of stimuli (i.e., List A, B, C, D, E, and F) were prepared for the speakers.

2.1.3. Noise Preparation

A controlled level of background noise was used in both of the production and the perception experiments. In the production task, the noise was fed through headphones to the speaker-participants to elicit the Lombard speech. In the perception experiment, some of the stimuli were digitally mixed with the same type of noise and presented with noise-free stimulus sentences to listeners (discussed in Chapter Three).

The masking noise was taken from a sound sample "Audience Walla A" available in the Earshot Sampler in the commercial software package SoundEdit 16 version 2.0.1. This noise was similar to cafeteria noise heard in a busy restaurant, as it seemed to consist of noises of cutlery, and unintelligible human speech. A short excerpt of the noise sample of five seconds was extracted with both end points at a zero crossing. This was then lengthened to a duration of about 30 minutes by concatenating at points of zero crossing using SoundEdit 16. The resulting file was recorded back onto tape using a JVC TD-W709 Cassette Deck. So that the noise could be presented binaurally to the talkers using the same tape deck via a pair of headphones during the formal recording section, the headphones were calibrated. The input recording level of the tape deck was manipulated, and the output level of the noise via the headphones was noted by monitoring the reading on a sound level meter (Bruel & Kjaer 2205). The headset was mounted on the sound level meter (SLM) that was fixed on a tripod with the internal surface inside the earmuff on each side lightly touching against the tip of the microphone of the SLM. The output level of the noise was monitored to ensure that the levels of noise from both sides of the headphones were equivalent. This method was also used in all the following measurements of the masking noise at various levels. A level of 70 dB (A) of the cafeteria-masking noise was selected, after the experimenter and a phonetician, using the same headset, had individually listened to the masking noise at various levels from 60 dB (A) to 90 dB (A) at 5-dB steps. It was verified that the 70 dB (A) level was neither too low to elicit the Lombard effect nor so high that it caused any discomfort to the ear. This was further confirmed by a linguistically-trained native speaker of English, naive to the purpose of the study, who went through the whole process in a trial recording. A similar type and level of noise (cafeteria noise at 70 dB (A)) was also used in an experiment about intelligibility of cellular phone speech in a noisy background conducted by Saito, Fukudome, and Tsumura (1999).

2.1.4. Recordings

Individual recordings were made in a sound-treated room in the Phonetics Laboratory of Simon Fraser University, where the level of background noise was found to be 25 dB (A) using the sound level meter (B&K 2205). Participants wore an MB Quart K800 headset equipped with a boom microphone, and their speech was recorded onto tape using a Marantz PMD 201 Cassette Recorder. As already mentioned in section 2.1.3., masking noise at 70 dB (A) was presented to the speakers through the headphones using the JVC TD-W709 Cassette Deck. One of the advantages of using this type of headset was that the microphone was kept at a constant distance (about 4 inches) from the lips of the participants throughout the recordings. Another advantage was that the recorded speech was noise-free because the masking noise fed through the headphones would not be captured by the boom microphone. Throughout the recordings, the speakers were comfortably seated face-to–face at a distance of two meters from the experimenter. The experimenter was present to simulate a communicative situation during the recordings in a laboratory environment. Although a number of studies have been conducted to look for differences between speech produced under stress and normal speech, it was noted by Junqua (1996) that these have been carried out in well-controlled conditions which are far-removed from real-world situations. In previous experiments like those done by Hanley and Steer (1949), Van Summers et al. (1988), Anglade and Junqua (1990), and Applebaum, Hanson, and Morin (1996), speakers were asked to read stimuli to themselves while listening to noise through headphones in the recording room. However, Junqua (1996) suggested that the magnitude of speakers' responses seem to be governed by the desire to produce intelligible communication with an interlocutor in the real world. In this experiment, the recording condition was made to be more realistic in a sense that the speakers listened carefully to instructions given by the experimenter who monitored what the speakers were saying throughout the recording sessions. Participants were requested not to start reading the first sentence on each page until the experimenter, who wore no headphone and did not listen to any background noise, said "start". They did not go on reading the next sentence until the experimenter said "next". Furthermore, the speakers were instructed to read the sentences in a conversation-like manner similar to the way they talked in everyday situations, as were the talkers in Picheny, Durlach, and Braida (1985, 1986). In the event of any reading errors, noticeable pauses or hesitations, the speaker was asked to repeat the sentence until a correct and fluent utterance was produced. Throughout the recordings, the frequencies of repetitions were less than 2.8% for the Cantonese speakers, and less than 1.2% for the English speakers. None of these repeated utterances was used in the subsequent perceptual experiment.

The participants read the 48 stimulus sentences on the test lists under two conditions: quiet and noise. As already mentioned, under the noise condition, the cafeteria-like masking noise was fed through the headphones at 70 dB (A). In contrast, no noise was presented in the quiet condition. The speakers were randomly assigned to read one of the six randomized sets of stimuli in each of the two conditions. For each speaker, no two sets were identical in both the experimental conditions. Three males and three females from each speaker group (i.e., a total of 12 speakers) were assigned to read the 48 stimulus sentences first in the quiet and then in the noise condition, while the others read the stimuli in reverse order of conditions. In the noise condition, speakers were given a 30-second break after reading a page. This was to prevent them from getting used to the masking noise (cf. Dreher & O'Neill, 1957), which might adversely affect the elicitation of the Lombard speech. In the quiet condition, no such restriction had to be imposed. A brief rest was given upon the speakers' request. A 5-minute break, nevertheless, was provided between recordings in the two conditions.

Prior to the recordings, the participants were given the list of 48 stimulus sentences, and instructed to read through it silently. So that speakers could give their best possible rendering of each sentence, they were permitted to practice reading through the whole list of statements immediately before the actual recording was made in each of the experimental conditions. It was also during this period that gain levels were adjusted as appropriate. In actual recording, the experimenter monitored the correct production of each sentence using a reading list monitoring sheet (see Appendix 4). Six versions of the sheet were prepared for each of the randomized lists.

2.1.5. Acoustical Measurements

A number of acoustical analyses were performed in an attempt to determine the differences between Lombard speech and normal speech. All utterances were digitally sampled at 22.05 kHz with a resolution of 16 bits using SoundEdit 16, and were saved as computer audio files. From each of the 24 speakers, four different productions (two true and two false) in each of the quiet and noisy conditions were selected for a total 192 utterances (12 speakers x 2 speaker groups x 4 sentences x 2 conditions). The criteria for selecting the particular sentences from each speaker will be discussed in Chapter Three. The four sentences chosen from each speaker in the quiet condition were identical to those in the noise condition.

In this experiment, two accent features, among those mentioned by Derwing and Munro (1997) and Munro (1995), were measured in every sentence: speaking rate and fundamental frequency. These two features, among others, were also used by Picheny et al. (1986) to examine acoustic differences in sentences spoken clearly and conversationally by native English speakers. To compute the speaking rate, the number of syllables in each of the 192 utterances was divided by the total speaking time for each sentence for each speaker (Anderson-Hsieh & Koehler, 1988; Derwing & Munro, 1997; Munro, 1995; Munro & Derwing, 1994, 1998, 1999). Duration for each sentence, measured to the nearest 0.01 second, was made from inspection of the time waveform combined with wide band spectrogram using the speech analysis software (Praat version 3.8.49).

Fundamental frequency (Fo) values were obtained so as to explore any differences in the sentence productions under the two different speaking conditions. First of all, measurements of the mean Fo values, the maximum and the minimum for each sentence, based on the waveform for duration measurement, were made with the PitchEditor in Praat using the autocorrelation algorithm. At this stage, any token with excessive vocal fry that resulted in a spurious Fo value was excluded so as to avoid difficulty for acoustic analysis (cf., Kehoe et al., 1995). The token in question would be replaced with another sentence produced by the same speaker. The analysis filter was set from 70- to 300-Hz for the male voices and 135- to 535-Hz for the female voices at every 40 msec (cf. Loren et al., 1986; Rivers & Rastatter, 1985). In their respective experiments, the researchers used the Kay Visi-pitch Fo analyzer (model 6087) in which the Fo range for male voice was fixed at 50- to 300 Hz. In a pilot study of this research, Li (1999) measured the mean sentential Fo values of 2 male Cantonese speakers and 2 male English speakers using that particular range (i.e., 50- to 300 Hz). However, he found that the male voice rarely dropped as low as 50 Hz in either the quiet or the noisy speaking conditions. Moreover, it has been suggested by Fry (1992: 68) that vocal fry (or creaky voice) is heard at the end of utterances where the fundamental frequency falls to a low level in the range of 20 to 60 Hz, and that it should not be taken into account in measurement of fundamental frequency. For this research, the Fo range for the male voices was thus set at 70- to 300 Hz. The Fo range for the female voice remained the same. In addition, the Fo Range for each sentence was also expressed and computed as a difference between the maximum Fo and the minimum Fo values (Reissland & Snow, 1996). Finally the relevant data were submitted to statistical analyses using statistical software (StatView version 5.0).

2.2. Results

2.2.1. Speaking Rate

The mean speaking rate (syll/s) of the four sentences (two true and two false) in each of the two speaking conditions was determined for each speaker. The overall mean speaking rates for the female and male speakers of the two groups are illustrated in Figure 2-1. The native English participants spoke at a faster rate (M = 6.41 syll/s for English females and M = 5.95 syll/s for English males) than did the native Cantonese speakers in the quiet condition (M = 4.50 syll/s for Cantonese females and M = 4.48 syll/s for Cantonese males). Similarly, the speaking rates for the Cantonese female speakers (M = 4.57 syll/s) and the Cantonese male speakers (M = 4.47 syll/s) were slower in the noisy condition than those for the English female speakers (M = 6.13 syll/s) and the English male speakers (M = 5.68 syll/s). The data were submitted to a mixed-design analysis of variance with Native Language of Speakers (NL), English and Cantonese, and Gender (GD), female and male, as between-subjects factors and Speaking Condition (SC), Quiet and Noise, as a within-subjects factor. The analysis yielded a significant effect of NL, F (1,20) = 21.15, $p < 0.01$, indicating that the English speakers did speak at a faster rate than the Cantonese speakers of English in the two speaking conditions. In contrast, neither the effect of GD nor that of SC approached significance, Fs (1,20) = 0.61 and 0.103, $ps > 0.05$. Overall, female speakers did not significantly differ from their male counterparts in speaking rates. In addition, the speakers did not speak at a significantly slower rate in the noisy condition than in the quiet condition.

On the other hand, there was a significant SC x NL interaction, F (1,20) = 4.85, $p < 0.05$, indicating that the speaking condition had an effect on the mean speaking rates of one of the speaker groups, but not the other. The native English talkers spoke at a slower rate in the noisy condition than in the quiet condition, but this phenomenon was not observed in the group of the Cantonese speakers of English. However, the two-way interaction (SC x GD), F (1,20) = 0.046, $p > 0.05$, and the three-way interaction (SC x NL x GD), F (1,20) = 0.076, $p > 0.05$, were found to be nonsignificant.

2.2.2. Sentential Fundamental Frequency Values

As already mentioned, four different productions (two true and two false) were selected from each speaker in each of the Quiet and Noise conditions. The mean, maximum, and minimum Fo values of each of the eight sentences for each speaker were measured, and the Fo range (maximum Fo – minimum Fo) for each utterance was also computed. Table 2-1 shows the mean Fo, maximum Fo, minimum Fo, and Fo range by gender and speaker group. The mean sentential Fo values were submitted to a mixed-design ANOVA with Native Language of Speakers (NL), and Gender (GD), as between-subjects factors, and Speaking Condition (SC), as a within-subjects factor. It is well known that regardless of speaker's native languages, the Fo value of female speech is generally higher than that of male speech (Fry, 1992: 68, Ladefoged, 1993: 187). The reason for still including Gender (female or male) as an independent variable here and in the following statistical analyses was to give a better overall picture of the observations and findings. This was similar to what Munro (1995) has done in describing the intonation patterns in English sentences produced by Mandarin and English speakers of both genders. Only the main effects for the three independent variables were found to be significant. First of all, results revealed a significant effect of the NL, F (1, 20) = 8.17, $p < 0.01$, indicating that the mean sentential Fo values for Cantonese speakers in Quiet (M = 227.93 Hz for females, M = 130.65 Hz for males) and in Noise (M = 252.37 Hz for females, M = 160.79 Hz for males) were significantly higher than those of English speakers in Quiet (M = 193.29 Hz for females, M = 121.43 Hz for males) and in Noise (M = 215.42 Hz for females, M = 148.82 Hz for males). Not surprisingly, a significant effect of GD was also found, F (1, 20) = 101.73, $p < 0.0001$. The analysis also revealed a significant main effect of the SC, F (1, 20) = 135.31, $p < 0.0001$. All speakers exhibited sentential mean Fo values that were higher in the Noise condition than in the Quiet condition. However, no significant interaction was found for SC x NL, SC x GD, or SC x NL x GD, Fs (1, 20) = 0.32, 1.50, 0.003, $ps > 0.05$. Figure 2-2 illustrates the mean sentential Fo values with respect to the native language, gender of the speakers, and the two speaking conditions.

Before the Fo range values of sentences were analyzed, the data for the mean minimum sentential Fo were submitted to another mixed-design ANOVA using the same factors as in the previous analysis. The purpose of statistically analyzing the mean values of minimum Fo was to form a basis for a later examination of the mean Fo range. As before, no significant interaction was found for SC x NL, SC x GD, or SC x NL x GD, Fs (1, 20) = 0.14, 2.36, 0.24, $ps > 0.05$. However, there were significant main effects on the mean minimum

Fo for NL, $F(1,20) = 4.91$, $p < 0.05$, and GD, $F(1,20) = 150.84$, $p < 0.0001$. Overall, the Cantonese speakers' mean minimum Fo values in Quiet ($M = 179.19$ Hz for females, $M = 99.28$ Hz for males) and in Noise ($M = 175.51$ Hz for females, $M = 108.72$ Hz for males) were significantly higher than those of English speakers in Quiet ($M = 155.75$ Hz for females, $M = 97.15$ Hz for males) and in Noise ($M = 157.62$ Hz for females, $M = 105.81$ Hz for males). However, no significant difference was found between the Quiet and Noise conditions overall, $F(1,20) = 1.58$, $p > 0.05$. The mean sentential Fo values with respect to the native language, gender, and the two speaking conditions are illustrated in Figure 2-3. It was clear that the mean minimum sentential Fo values for the female and male talkers in both groups were more or less the same in the two conditions. In other words, the Lombard reflex did not have a significant effect on the minimum sentential Fo value for all speakers.

Figure 2-4 illustrates the mean values of the sentential Fo range (maximum Fo – minimum Fo) with respect to the native language and gender of the speakers under the two experimental conditions. The computed data for the mean sentential Fo range were submitted to another mixed design ANOVA using the same factors as before. It yielded a significant effect of GD, $F(1,20) = 21.18$, $p < 0.01$, as well as a significant effect of SC, $F(1,20) =$

173.82, $p < 0.0001$. However, the effect of NL failed to reach significance, F (1,20) = 1.80, $p > 0.05$. All of the speakers, regardless of their native language, exhibited a wider sentential Fo range in Noise than in Quiet. In addition, neither the interaction of SC x NL nor that of SC x NL x GD was found to be significant, Fs (1,20) = 0.25 and 0.59, $ps > 0.05$. However, there was a significant interaction between SC and GD, F (1,20) = 4.62, $p < 0.0441$. It was probably due to the wider Fo range exhibited by the female speakers than by the male speakers.

2.3. Discussion

2.3.1. Speaking Rate

First of all, it was noted that the rates of speech for all speakers in this experiment were faster than those for speakers observed in previous research. In a recent study by Munro & Derwing (1999) using similar true and false sentences as reading materials, for instance, the mean speaking rates for 10 Mandarin speakers of English (5 female, 5 male) and for 4 English speakers (2 female, 2 male) were 3.52 and 4.55 syll/s, respectively. Upon comparison, it can be readily seen that even the mean speech rate for the native Cantonese speakers in this study was comparable to the rate for their English speakers. This is probably because all speakers in this experiment were instructed to read in a conversation-like manner. As indicated by Picheny et al. (1986), speaking rate for conversational speech is generally faster than that for natural speech. The instruction given to the subjects here appears to be an effective and helpful way of eliciting speech that is more natural than speech recorded in a simple reading task.

As in previous studies, the native English speakers spoke at a significantly faster rate than did the native Cantonese speakers in both the speaking conditions. This should not be surprising, as Munro & Derwing (1998) have already noted that nonnative speakers typically speak slower than do native English speakers. The slower-produced English sentences may be due in part to the different rhythmic patterns of L1 and L2 of the native speakers of Cantonese (or Chinese, in general). Roach (1982) suggested that syllable-timed languages, such as Cantonese (Bauer & Benedict, 1997), have simpler syllable structure, but that stress-timed languages, such as English, exhibit more vowel reduction in unstressed syllables. As also mentioned by Roach (1982), it is commonly believed that for syllable-timed languages, all syllables will be given equal amounts of time; while for stress-timed languages, more time is allotted to stressed and less to unstressed syllables. He (1987) observed that Chinese learners of English employ the speech rhythm of Mandarin in producing English sentences: marking stress on every syllable, and attempting to articulate every word (content and function) as clear as possible. It is possible that the Cantonese speakers for this study employed the same strategy. Given the lack of pauses or hesitations in a sentence-length utterance, it is not difficult to understand why Chinese (or Cantonese) speakers' speech rates will be relatively slower than those of native speakers of English. In addition, the differences in speaking rates may be due to longer vowels and sonorants in sentences produced by nonnative speakers, as found by Guion et al. (2000).

In addition, no significant difference was found in the rate of speech between the female and male speakers for both groups in each of the two speaking conditions. This outcome differs from Byrd's (1992) finding that female speakers of American English speak more slowly than their male counterparts. This discrepancy in the effect of gender on speaking rate may be partly due to the great difference in the number of speakers employed in the experiments. From the TIMIT speech database, Byrd (1992) used 420 American English talkers of whom 31% were female, while this study only had 12 speakers of each gender (pooled over the two groups of speakers). Since the results reported here are only based on such a small number of speakers, differences in speaking rate between women and men may not have been observed. Another possible factor may be related to the inclusion of nonnative speakers of English (i.e., Cantonese speakers) in this study, the gender effect of which on speaking rate in producing L2 sentences has not yet been fully known.

Surprisingly, the masking noise used in this experiment did not have a significant effect in slowing down the speaking rates for the speakers overall. The result was contradictory to what Hanley and Steer (1949) had found. A possible explanation for the discrepancy may be due in part to the type of noise used for masking purpose. Junqua (1996) has pointed out that various types of noise may have different effects in shaping the acoustic changes in the speech signals. As already mentioned, the type of noise used in Hanley and Steer (1949) was aircraft noise, while the one used here was cafeteria-like with noises of cutlery and unintelligible human speech. It is possible, then, that unlike the aircraft noise, the cafeteria-like masking noise does not have a significant influence in reducing the speaking rates for all of the speakers in this experiment.

However, it was noteworthy that the native English speakers, but not the Cantonese speakers, did speak more slowly in the Noise condition than in the Quiet condition.[1] There seem to be a number of plausible reasons for the Cantonese speakers not slowing their speech rate in noisy condition. First, it is known that L2 speakers of English already speak significantly slower than do native English speakers in normal recording situations (Munro, 1995; Munro & Derwing, 1998, 1999). It is possible that these nonnative English speakers, when speaking under noisy conditions, cannot naturally or unconsciously reduce even further their speaking rates, unless they are explicitly instructed to do so (e.g., Munro & Derwing, 1998). Moreover, Hanley and Steer (1949) have suggested that even native speakers who have not been trained or instructed know how to react appropriately in an adverse communicative situation by reducing their speaking rates. For the native Cantonese speakers, English is their second language. It is not unreasonable to propose

---

[1] Since it has been reported that vowel duration increases in utterances produced in noisy conditions, the vowel durations in the word "cats" in sentence no. 27 (see Appendix 3) produced by the speakers in quiet and in noisy conditions were measured. The segmentation followed the criteria suggested by Peterson and Lehiste (1960). After excluding the measurement of one outlying data point (i.e., more than two standard deviations from the mean), the data were submitted to a mixed-design ANOVA with NL as a between-subjects factor and SC as a within-subjects factor. For this particular vowel, there was neither a significant effect of NL nor a significant interaction of NL x SC, $Fs\ (1, 21) = 3.05$ and 0.12, $ps > 0.05$. In contrast, the effect of SC on vowel duration was found to be significant, $F\ (1, 21) = 11.38$, $p < 0.01$, indicating that the vowel duration increased in Lombard speech for both groups of speakers.

that they simply do not know how to adjust their speech aptly to produce the desired Lombard effect in English. It should also be mentioned that the "resistive" performance of the native Cantonese speakers also brings out an interesting issue. As far as the properties of the Lombard reflex are concerned, the effects of the background noise on L1 may not necessarily be felt to the same extent as on L2. In general, the type of the masking noise and the native language investigated in the experiment may all affect the changes observed in the speech signals, thus complicating the subject matter to a greater extent (Junqua, 1996).

2.3.2. Fundamental Frequency

For the mean sentential fundamental frequency, as was the case for the Mandarin speakers of English in Munro (1995) and of the Japanese speakers of English in Yamazawa and Hollien (1992), the Cantonese speakers exhibited higher mean Fo value than did the native English speakers in producing the stimulus sentences under all speaking conditions. In addition, the minimum Fo values for the Cantonese speakers were also higher than those for their native counterparts. The reasons for the Cantonese speakers' utilizing an overall higher fundamental frequency than the English speakers are not fully understood. As already suggested by Yamazawa and Hollien

(1992), factors such as height and weight are unlikely to be highly correlated with fundamental frequency, and thus, are not considered as either selection criteria or experimental variables. However, it is possible that the difference may rest on the structural characteristics of the speakers' native languages. As is the case for Japanese, a syllable-timed pitch accent language (Yamazawa & Hollien, 1992), the tone aspect of Cantonese may account for the addition of higher frequencies that leads to the higher mean Fo relative to that of the native English speakers. Fry (1968: 368) has already noticed that the Cantonese tones are less modified by intonation in Cantonese sentence. If this is the case, it is likely that some carryover from L1 for the Cantonese speakers influences their English productions in such a way that sentential Fo is even higher than for native English speakers. Nevertheless, this language-dependent phenomenon needs further research.

In addition, the results are consistent with earlier research finding (Bond et al., 1989; Loren et al., 1986; Rastatter & Rivers, 1983; Van Summers et al., 1988) showing that, as expected, the mean Fo significantly increased when the speakers spoke during exposure to noise. It had been reported by Lane and Tranel (1971) that an increase in fundamental frequency, among other variables, is a component of the Lombard reflex, despite the belief that this feature is somewhat more variable and less sensitive than a change in intensity. In fact, pitch change is a normal accompaniment of a change in vocal intensity, which is probably associated with changes in subglottal pressure and/or in vocal cord tension (Lane & Tranel, 1971).

In contrast, the mean minimum Fo values for all speakers were similar in the Quiet and the Noise conditions. It appears, then, that the Lombard reflex has no significant effect in raising the minimum sentential Fo values for the speakers. It is highly likely that in both the speaking conditions, the speakers tended to give an end-of-sentence signal (Picheny et al., 1986). Normally, the minimum fundamental frequency is usually observed near the end of a simple declarative sentence in English (Cooper & Sorensen, 1981: 3).

It was interesting to note that regardless of whether the subjects are L1 or L2 speakers of English, the ambient noise did have a greater effect of extending their Fo ranges. Van Summers et al. (1988) stated that the acoustic-phonetic changes associated with the Lombard effect are similar to those observed in clear speech, of which a wider Fo range is one of the characteristics, as noted by Picheny et al. (1986). Since no significant difference has been found for the minimum Fo between the two speaking conditions, the increased Fo ranges implies a greater laryngeal tension that results in greater peak Fo values under the influence of the Lombard reflex.

2.4. Conclusion

The primary concern of this experiment was to explore the effects of the Lombard reflex on sentences spoken by the nonnative speakers of English. This was done by examining and comparing differences for the two accent features, speaking rate and fundamental frequency values, in English sentences produced in noise-free and in noisy environments with cafeteria-like masking noise at a level of 70 dB (A) being fed through the headphones to the native Cantonese and the native English speakers. For speaking rate, the native English subjects spoke at a faster rate than did the Cantonese speakers of English in the two speaking conditions. Also, the female speakers' speech rate was more or less the same as the male counterparts in quiet and in noisy conditions. Most unexpected was the finding that the masking noise seemed to have no significant effect on slowing down the speech spoken by the speakers overall. Only the native English speakers, but not their nonnative counterparts, spoke slowly in the noisy condition. In contrast, all of the speakers exhibited a higher sentential Fo and a wider Fo range in the background of masking noise than in the noise-free environment. In addition, the fundamental frequency values for the Cantonese speakers were greater than those for the native English speakers not only in the quiet, but also in the noisy conditions.

It appears that the Lombard reflex is dependent on parameters such as the type of masking noise and the characteristics of the native languages of the speakers (Junqua, 1996). Furthermore, a noisy environment may affect the auditory feedback that results in a modification of normal speech production (Junqua, 1996). As a result, the changes in features of the speech signals (i.e., the Lombard speech) reflect articulatory modifications made to increase the vocal effort and to articulate more precisely for more intelligible communication in an adverse listening condition (Lane & Tranel, 1971). Generally speaking, the present study confirmed the changes accompanying the Lombard reflex for the speech signals in terms of one (i.e., Fo values), but not both, of the two accent features in this experiment. This carries an implication that the effects of the Lombard reflex on a speaker's first language may not necessarily be equivalent to those on speaker's second language. This experiment can serve as a departure point for future research examining any different effects of noise on first and second languages for bilingual speakers.

Junqua (1996) noted that acoustic differences between normal speech and Lombard speech contribute to the intelligibility of speech, and thus speech produced in noise is more intelligible than speech produced in a quiet condition. This proposal is explored in Chapter 3 where the intelligibility, as well as comprehensibility and accentedness, of normal speech and Lombard speech are examined in a perception experiment. Some of the productions from each of the two speaking conditions for each speaker will be mixed with the same type of making noise at a constant signal-to-noise ratio and presented, with noise-free tokens, to a new group of English native-speaker listeners.

TABLE 2-1

Mean values (Hz) of the Fo, maximum Fo (Max), minimum Fo (Min), Fo range (maximum Fo – minimum Fo) (Range), and standard deviations (in parentheses) by gender and group of the speakers under the Quiet and Noise conditions.

Gender / Speaker Group		Condition: Quiet Mean	Quiet Max	Quiet Min	Quiet Range	Noise Mean	Noise Max	Noise Min	Noise Range
Female									
Cantonese	Mean	227.93	284.25	179.19	105.06	252.37	333.34	175.51	157.83
	(S.D.)	(21.71)	(24.60)	(18.87)	(20.26)	(20.46)	(23.11)	(7.93)	(27.24)
English	Mean	193.29	250.66	155.75	94.90	215.42	294.60	157.62	136.98
	(S.D.)	(15.37)	(28.11)	(14.14)	(30.79)	(21.04)	(38.83)	(14.93)	(35.68)
Male									
Cantonese	Mean	130.65	163.25	99.28	63.97	160.79	208.79	108.72	100.08
	(S.D.)	(25.13)	(32.88)	(19.97)	(29.76)	(26.66)	(34.12)	(12.27)	(26.66)
English	Mean	121.43	150.70	97.15	53.55	148.82	191.51	105.81	85.70
	(S.D.)	(14.18)	(20.12)	(8.13)	(15.47)	(16.94)	(23.74)	(18.99)	(21.07)

FIGURE 2-1

Mean speaking rates according to the native language and gender of speakers under the two speaking conditions (Quiet and Noise).

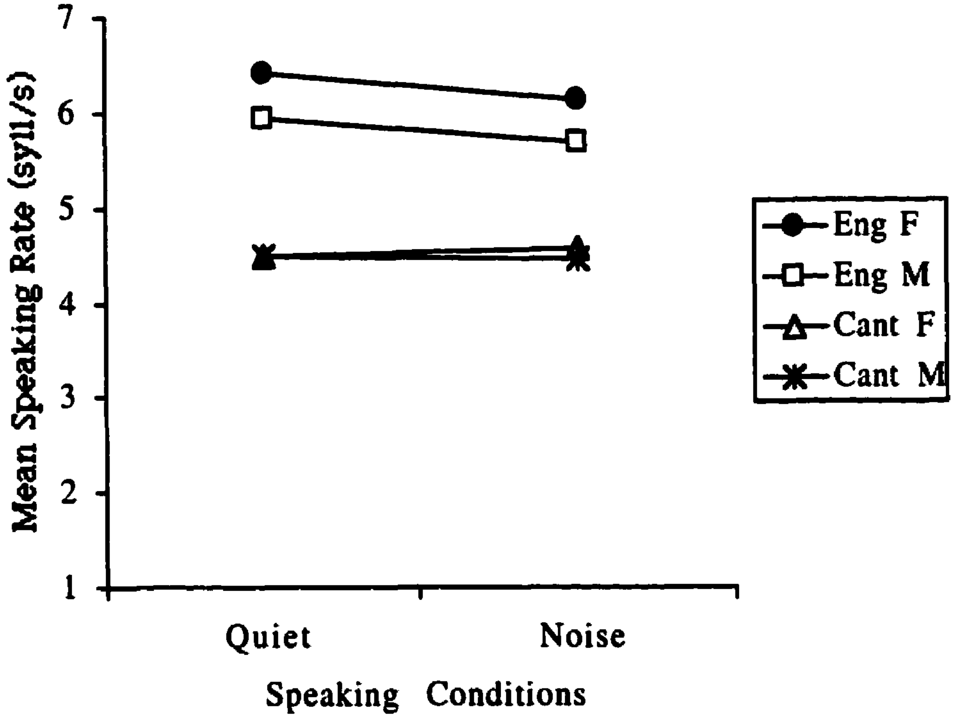


FIGURE 2-2

Mean sentential fundamental frequency values (Hz) according to the native language and gender of speakers under the two speaking conditions (Quiet and Noise).

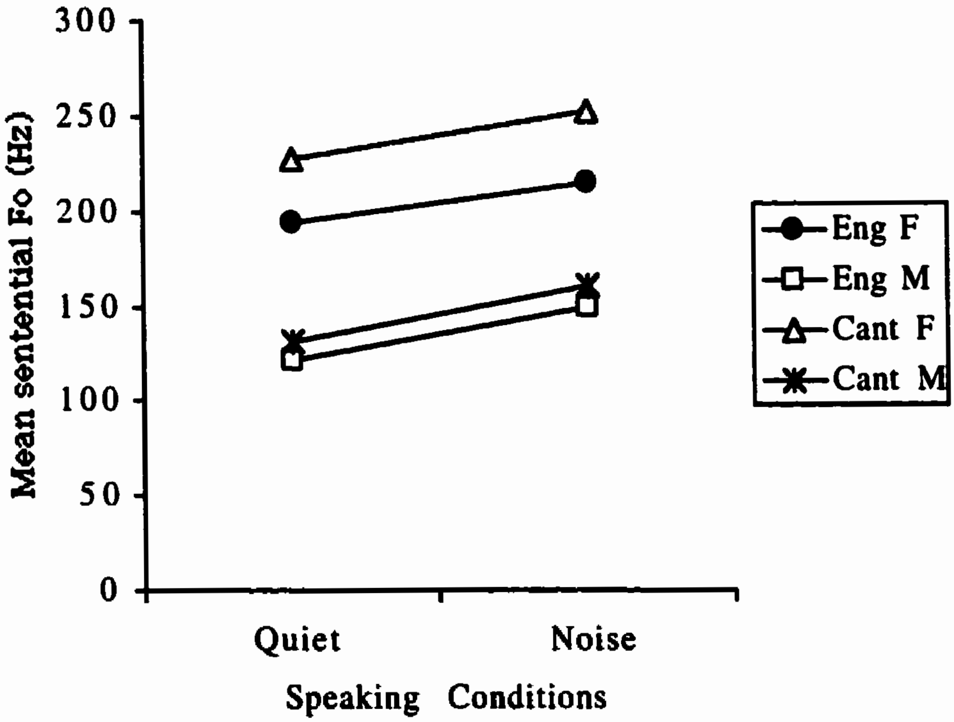


FIGURE 2-3

Mean sentential minimum Fo (Hz) according to the native language and gender of speakers under the two speaking conditions (Quiet and Noise).

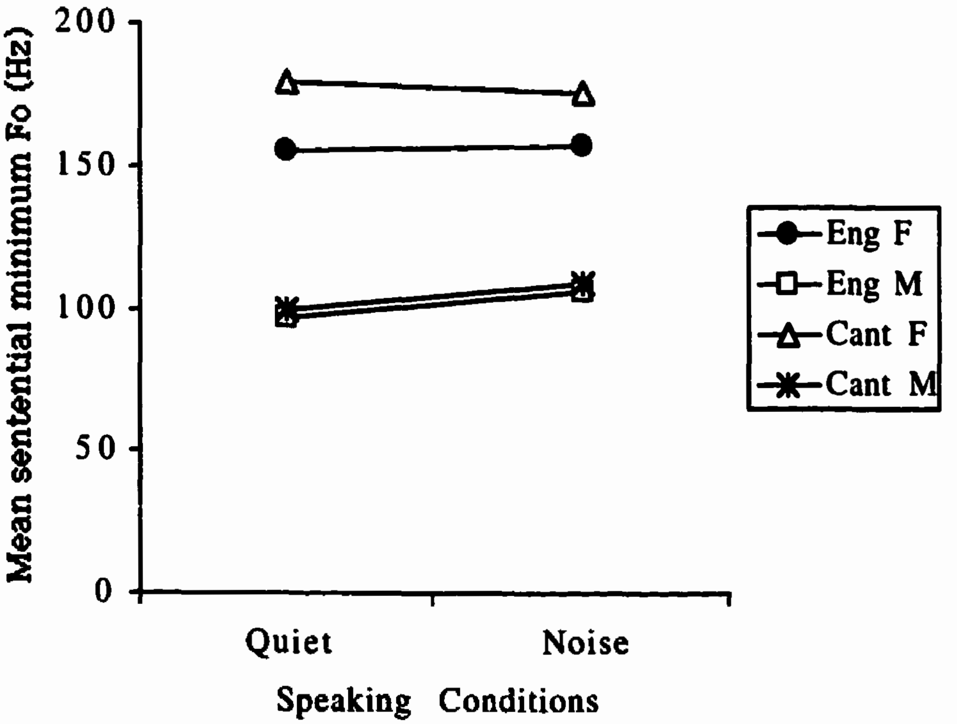
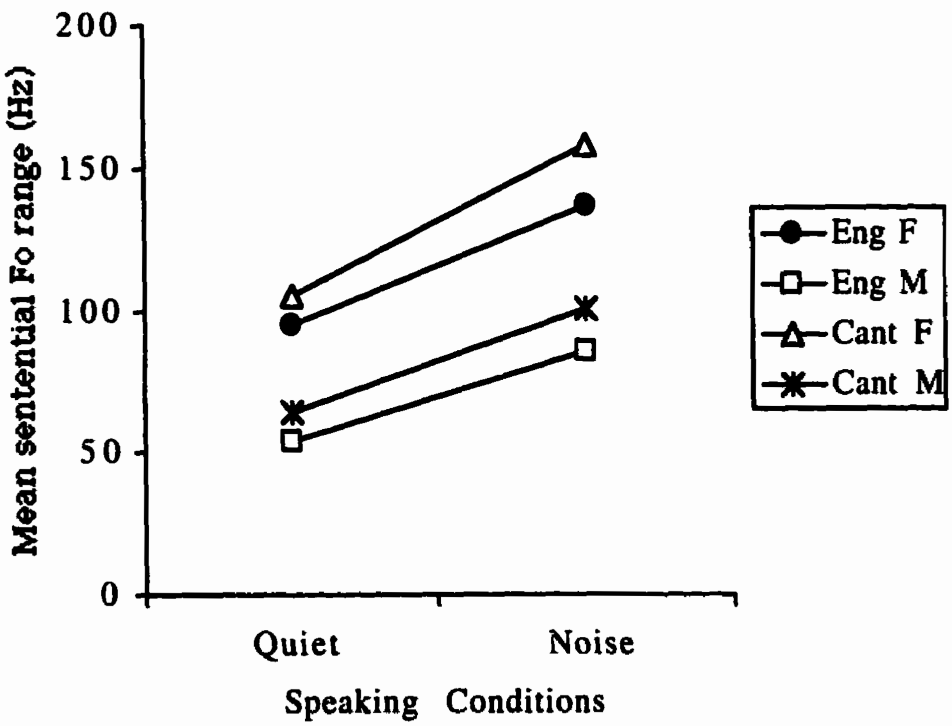


FIGURE 2-4

Mean values of sentential Fo range (Hz) according to the native language and gender of speakers under the two speaking conditions (Quiet and Noise).



CHAPTER III

EXPERIMENT 2: EFFECTS OF NOISE ON PERCEPTION OF CANTONESE SPEAKERS' ENGLISH SENTENCES

The purpose of this experiment is to explore the effects of masking noise on the perception of Lombard and normal speech produced by nonnative speakers of English, as well as by native English speakers. As already mentioned, little research on the production of Lombard speech has been done with nonnative speakers. Furthermore, studies of the effects of noise on the perception of nonnative Lombard speech have been seriously lacking.

In the following experiment, a new group of native English-speaking listeners (8 female) were presented with a list of English sentences produced by native speakers of Cantonese and of English in quiet and in noisy environments. The simple true and false sentences were selected from productions in Experiment 1. Some of the Lombard and normal speech samples were mixed with the same type of masking noise used in the production task, and were presented to listeners with unmasked sentences, all equated with root-mean-square amplitude. The listeners assessed the intelligibility of the speech in a sentence-transcription task and a sentence-verification task. In addition, the listeners evaluated the degree of foreign-accentedness and comprehensibility of the sentence-length utterances by assigning ratings on a 9-point scale.

It has been reported in the literature concerning L2 speech that nonnative English is often rated as more accented, less intelligible and less comprehensible than native-produced English (Anderson-Hsieh & Koehler, 1988; Derwing & Munro, 1997; Derwing, Munro & Wiebe, 1998; Munro, 1995, 1998; Munro & Derwing, 1994, 1995a, 1995b, 1998, 1999). It is expected that in this study, the speech produced by the Cantonese speakers is perceived to be less intelligible, more accented and more difficult to understand than that produced by the native English speakers in all listening conditions.

Speech presented without noise is generally reported to be more intelligible than that presented in a noisy environment (Dreher & O'Neill, 1957; Lane, 1963; Munro, 1998). As a result, it is expected that in the present study, the sentences presented in the noise-free condition will be more intelligible and more comprehensible than those presented with masking noise. As suggested in previous studies (Dreher & O'Neill, 1957; Van Summers et al., 1988), when either words or sentences are presented in noise at a constant signal-to-noise ratio, the speech originally produced in noisy conditions is more intelligible than speech produced in quiet conditions. For all the speakers in the present experiment, it is not unreasonable to posit that Lombard speech is more intelligible and more comprehensible than normal speech when both types of speech are presented to listeners in quiet and in noisy environments.

In addition, as noted earlier by Munro (1998), foreign-accented speech is harder to understand than is native-produced speech in noise. It is anticipated that the masking noise will decrease the intelligibility and perceived comprehensibility of the nonnative-produced speech to a greater extent than those of the native speakers. However, it is not yet known to what extent masking noise influences the degree of foreign-accentedness in normal and Lombard speech produced by the Cantonese speakers of English. This will also be explored in the current study.

3.1. METHOD

3.1.1. Listeners

A new group of native speakers of Canadian English (8 females) served as listeners. They were undergraduate students at

Simon Fraser University who had been born and raised in Vancouver, British Columbia. Their ages ranged from 19 to 29 years (M = 21.5 years), and all passed a hearing screen (250-4000 Hz at 20 dB) prior to performing the listening tasks. None spoke Cantonese or had any prior knowledge of the list of statements to be used for the experiment. They completed a Language Background Questionnaire similar to the one filled out by the native English speakers. None had previously participated in any speech experiment of a similar nature. They were paid an honorarium of $15 upon completion of the experiment.

3.1.2. Stimulus Preparation

The following sections describe the preparation of the stimuli for the perceptual experiment. The preparation consisted of three stages. In the first stage, overall root-mean-square (RMS) amplitudes of all the stimuli were determined and equalized. During the second, a number of pilot experiments were carried out to determine the appropriate signal-to-noise (S/N) ratio for the stimulus sentences. Creation of different sets of stimuli for presentation to the listeners was completed in the last stage.

3.1.2.1. Amplitude Equalization

The true and false sentences produced by the speakers in the production experiment were used as stimuli for the perception task. The truth value of all the statements could be easily determined by individuals on the basis of general knowledge (e.g., “A tiger is bigger than a cat”, or “You can tell time with a kettle”). To create the full stimulus set, four different sentences (two true and two false) produced in each of the quiet and noisy speaking conditions were selected from each of the 24 speakers for a total of 192 utterances (12 speakers x 2 speaker groups x 4 sentences x 2 speaking conditions). From each speaker, the four specific sentences produced in the quiet condition (i.e., Q-N) were identical to those in the noisy condition (i.e., L-N). The four pairs of statements (Quiet versus Lombard) were randomly selected with the requirement that the complete set of stimuli contained two pairs (Quiet versus Lombard) of each sentence from the list. For example, a pair of sentence No. 1 tokens (Q-N and L-N) was selected from speaker No. 1, while another pair of sentence No. 1 tokens (Q-N and L-N again) was chosen from another speaker, say, speaker No. 2. These ultimately constituted four complete lists of 48 stimulus sentences with two of them consisting of Quiet speech (Q-N) and two consisting of Lombard speech (L-N). Within each list, every speaker was represented by one true and one false statement. The four lists were paired into two groups (Group One and Group Two), such that each group was formed with a list of Quiet speech (Q-N) items and a list of Lombard speech (L-N) items having identical specific sentences from the same speakers (see Appendix 5). Q-N_1 and L-N_1 signified the lists in Group One, and Q-N_2 and L-N_2 represented those in Group Two.

To ensure a constant signal-to-noise (S/N) ratio across all stimuli (Munro, 1998; Van Summers et. al., 1988), the root-mean-square (RMS) amplitude for each of the 192 statements was determined and equated. The RMS amplitude value (in dB) was obtained using the RMS Envelope algorithm of Signalyze (Keller, 1994), according to the following formula (Keller, 1994: 80).

$$\text{dB} = 20 \times \log_{10} (\text{raw RMS value}/1)$$

The second columns in both Appendix 6 and Appendix 7 give the individual RMS amplitude values (in dB) of all the sentences and the overall means in Q-N_1 and L-N_1, while the ones in Appendix 8 and Appendix 9 show those values for all the utterances in Q-N_2 and L-N_2. It was found that in Group One, the mean of the RMS amplitude values for all the sentences in Q-N_1 was 50.08 dB, while the mean for L-N_1 was 50.10 dB. In Group Two, the means of the RMS amplitude values for all the utterances were found to be 50.98 dB in Q-N_2 and 49.93 dB in L-N_2.

On the basis of the computed RMS amplitudes, it seemed reasonable to equate the RMS amplitudes for all the stimulus sentences to a level of 50 dB before adding the masking noise. The value of 50 dB was considered appropriate after many pilot trials. When a speech signal with a greater RMS value was mixed with the masking noise of a similar level, the combined signal could show clipping when using Signalyze. This phenomenon is also documented by Keller (1994). The amplitude adjustments were accomplished by first calculating the raw RMS amplitude value for each of the 192 sentences on the basis of its dB level. An attenuating factor (AF) expressed in percent was then computed as a result of comparing the raw RMS value for each of the sentences to that of the target level (i.e., the raw RMS value of the 50 dB standard). Finally, the speech signals were multiplied by their respective AF in the amplification algorithm of SoundEdit 16 so that they all had the same mean RMS amplitude. The computation was performed using the formula on p.59.

For illustrative purpose, Y here represents the RMS amplitude value, and R_Y denotes the raw RMS value at Y dB for any sentence.

$$Y\ dB = \log_{10} (R_Y /1) \times 20$$

$$\log_{10} (R_Y /1) = Y/ 20$$

$$R_Y /1 = \text{antilog}_{10} (Y/ 20)$$

$$R_Y = 10^{(Y/ 20)}$$

Since the target level is set at the value of 50 dB, the calculation of its raw RMS amplitude value, represented by R_{50}, is as follows:

$$50\ dB = \log_{10}(R_{50} /1) \times 20$$

$$\log_{10} (R_{50} /1) = 50/ 20$$

$$R_{50} /1 = \text{antilog}_{10} (50/ 20)$$

$$R_{50} = 10^{(50/ 20)}$$

$$R_{50} = 316.23$$

To compute the AF in percent,

$$AF = (R_{50} / R_Y) \times 100\%$$

$$AF = (316.23/ R_Y)\ \%$$

The last three columns in Appendix 6 to Appendix 9 show the raw RMS value, AF, and AF rounded to the nearest 1% for all the sentences in the two Groups. The overall RMS amplitudes (dB) for all the sentences were modified to the desired 50 dB level by entering the respective AF (%) in the "Amplify" effect of SoundEdit 16. As a precaution, the RMS (dB) values of all the processed sentences were measured again using the RMS envelope algorithm of Signalyze. It was verified that the RMS (dB) values for all the sentences were now more or less the same as the target level of 50 dB. The percentage difference between the maximum and the minimum RMS (dB) values of sentences within each of the lists was less than 0.4 %. This error, which was deemed tolerable for the purposes of this study, was due to rounding of the AF (%). Appendix 10 and Appendix 11 show the maximum, minimum, and mean RMS (dB) values with standard deviations for all the processed sentences in Q-N_1 and L-N_1, while Appendix 12 and Appendix 13 give those in Q-N_2 and L-N_2. One token of each sentence was kept in its original noise-free form, and a duplicate token of each statement was prepared for mixing with the masking noise. Therefore, the stimulus set now contained a grand total of 384 sentences. Within each group, there were four complete lists of 48 statements, with two of them being normal speech and the other two Lombard speech (see Appendix 5). Q-N and L-N were referred to as clean speech, whereas the duplicate ones with added noise, Q+N and L+N, were referred to as noisy speech.

3.1.2.2. Piloting

In this research, three pilot perceptual experiments were carried out for two different purposes. Before the masking noise was digitally mixed with the speech signals for the formal perceptual experiment, an appropriate level of noise had to be determined. The proper level was selected with the underlying rationale that the noisy sentences would neither be too easy nor too difficult for the listeners to verify. In the first two piloting sessions, a S/N ratio was established that would lead to a moderate amount of degradation in which between 70% and 80% of the noisy sentences still remained partially or completely intelligible at the selected level of the masking noise (Munro, 1998). Another trial listening test was also conducted to ensure the intelligibility of the clean Lombard utterances produced by the native Cantonese and the native English speakers. This preliminary examination of the clean Lombard speech seemed to be necessary because the Lombard speech, rather than the normal speech, was the main focus of the research. Any subsequent analysis of listeners' scores on the noisy Lombard speech in the actual experiment would be pointless if the clean Lombard speech was unintelligible at the outset.

So that the stimuli for the first two pilot studies contained both normal speech and the Lombard speech to be mixed with noise, a full list of 48 trial utterances was made up of sentences selected from both Q-N_1 and L-N_2. Twenty-four utterances were from Q-N_1 and the rest from L-N_2. Of the 24 sentences from each of the lists, half were true utterances and half were false with a balance in the speakers' L1. The specific sentences from each speaker were the same for these two pilot experiments. For the third pilot experiment, sentences in the two lists of L-N_1 and L-N_2 were used instead. All sentences were saved as computer audio files so that they could be randomly presented to the listeners through a custom response-collection computer program (Munro, 1997).

The masking noise added to the sentences was extracted from a five-second-long sample, the amplitude of which was adjusted in accordance with the trial S/N ratios using the computation in Section 3.1.2.1. The mixing of the speech signals with the masking noise at any particular noise level was done using SoundEdit 16. Details will be discussed later in this section.

A sentence-verification task was used to assess the intelligibility of the sentences throughout the three pilot studies (Munro, 1998). Details are given in section 3.1.3, as this task was also used in the formal perceptual experiment. Briefly, upon listening to the stimulus sentence once, listeners in all the three pilot tests had to determine the truth value of the sentence by selecting one of three buttons ("True", "False", or "Unknown") on the computer screen. The verification scores were based on the number of correct true and false responses (Munro, 1998).

The first piloting session was used to evaluate the intelligibility of the noisy sentences at three different S/N ratios: -5, -3, and 0 dB. The RMS amplitudes (dB) of the noise levels were 55 dB, 53 dB, and 50 dB, respectively, with the speech signals being fixed at the target level of 50 dB. A complete list of 48 noisy sentences was used as stimuli. For each S/N ratio, 16 sentences were mixed with the appropriate masking noise. Within the 16 sentences, four true and four false utterances (i.e., a total of eight) were selected from each of the Q-N_1 and L-N_2 sets.

Two female volunteers participated in this trial experiment. One of them was a 42-year-old native speaker of Mandarin Chinese, and the other a 30-year-old native speaker of South Korean. Both of them were graduate students in the Department of Linguistics of Simon Fraser University. They passed a binaural pure-tone hearing screen (250, 500, 1000, 2000, and 4000 Hz at 20 dB) prior to performing the perceptual experiment.

It was found that the mean percentage score (pooled over the two listeners) for the verification task at 0 dB S/N ratio was 71.88%. The score was 37.50% at –3 dB S/N ratio, and it dropped to 28.13% at –5 dB S/N ratio.

The above results revealed that the S/N ratio at 0 dB appeared to be a potentially suitable final noise level. However, the first pilot session was still considered a rather coarse estimation, because it had covered a wide range of S/N ratios. In the second pilot perceptual experiment, we refined the approximation by narrowing the range of S/N ratios from –1 to +1 dB with 0 dB as the middle value. The RMS amplitudes (dB) of the noise levels were 51 dB, 50 dB, and 49 dB, respectively, and those of the speech signals were kept at 50 dB. As in the first pilot session, 16 sentences were mixed with the appropriate level of the masking noise at each S/N ratio.

Four native English speakers (one male, three female) served as participants in this trial experiment. All the listeners were speakers of Canadian English who had been born and raised in British Columbia. They were all undergraduate students at Simon

Fraser University with a mean age of 21.3 years (range: 20 to 22 years). All of them passed the pure-tone hearing screen. Again, none had participated in any speech experiment involving noise.

The mean percentage score (pooled over the four listeners) for the verification task at -1 dB S/N ratio was 54.69%. The values at 0 dB and +1 dB S/N ratios were 68.75% and 78.13%, respectively. On the basis of these results, it was decided that in order to meet the target of 70 – 80% correct verifications, the S/N ratio should be set at +1 dB. In other words, the RMS amplitude of the noise should be at a level of 49 dB with that of the speech sample at 50 dB.

The third pilot listening task concerned the intelligibility of the clean Lombard speech. Two new female native English speakers voluntarily participated in this trial session. Both were graduate students in the Department of Linguistics of Simon Fraser University. One of them was 50 years, and the other was 52 years. Both had passed the pure-tone hearing screen. None had participated in any speech experiment that involved noise. They were randomly assigned to listen to one of the lists (i.e., clean L-N_1 or L-N_2) and verified the truth value of all the stimulus sentences. As already mentioned, within each list, half of the 48 sentences were produced by the native English speakers, while the rest were produced by the Cantonese speakers of English.

It was found that for the clean Lombard speech produced by the native English speakers in the two lists, the overall correct verification score (pooled over the two listeners) was 97.92%. One of the listeners failed to correctly verify one (out of 24) sentence produced by a male speaker in the L-N_2 list, while the other listener obtained 100% correct verification on list L-N_1. In contrast, the verification score for the speech produced by the native Cantonese speakers was only 79.17% overall. After the task, the listener who made the wrong verification was asked why she might have incorrectly verified one sentence. She stated that her error was due to the choppy intonation pattern exhibited in that false sentence. However, she made no such judgement on the true sentence produced by the same English speaker within the same stimulus list. From the results of this pilot experiment, the degree of intelligibility of the clean Lombard speech produced by the native English speakers could be considered acceptable for this research.

As mentioned before, the two lists Q-N and L-N in each group (Group One and Group Two) had been duplicated so that the stimulus sentences could be mixed with the masking noise. The mixing was carried out using the "Mix" effect of SoundEdit 16. In Figure 3-1, the top panel is the speech waveform of one of the 48 sentences, which was spoken by a female Cantonese speaker in quiet conditions. The middle panel shows the masking noise. The noisy sentence, a combination of the speech signal and the masking noise, is shown in the bottom panel. As in Hazan & Simpson (1998) and Van Summers (1988), every noisy token was preceded and followed by a segment of the masking noise of 100 msec.

After the mixing process, all the noisy sentences were again equated in terms of overall RMS amplitude at the same level as that of the clean speech (i.e., 50 dB). The two lists of noisy sentences were denoted as Q+N and L+N, respectively. As a result, Group One now contained four lists of 48 sentences ($Q\text{-}N_1$, $Q+N_1$, $L\text{-}N_1$, and $L+N_1$). Four other lists ($Q\text{-}N_2$, $Q+N_2$, $L\text{-}N_2$, and $L+N_2$) belonged to Group Two. As a precaution again, the RMS amplitude (dB) values of all the noisy sentences were measured. It was found that the RMS amplitude (dB) values for all the noisy sentences were very close to the target level of 50 dB with slight variations due to rounding error. The percentage differences between the maximum and the minimum RMS (dB) values of these sentences within each of the lists were less than 0.4 %. Appendix 14 and Appendix 15 present the maximum, minimum, and mean RMS (dB) values with standard deviations for all the noisy sentences in $Q+N_1$ and $L+N_1$, while Appendix 16 and

Appendix 17 give those in $Q+N_2$ and $L+N_2$. It should be mentioned that in each group, the specific true and false sentences spoken by any speaker were the same throughout the four speech conditions (i.e., Q-N, Q+N, L-N, and L+N) (see Appendix 5).

3.1.2.3. Creation of stimulus sets

Four separate stimulus sets were prepared according to the following criteria. Each set consisted of the full complement of 48 test items (24 true, 24 false). Individual speakers were represented twice, once producing a true sentence and once producing a false sentence. The sentences for two of the sets (Set 1A and Set 1B) were selected from those in the four lists of stimuli (i.e., $Q-N_1$, $Q+N_1$, $L-N_1$, and $L+N_1$) of Group One. For the other two sets (Set 2A and Set 2B), the sentences were selected from the four lists (i.e., $Q-N_2$, $Q+N_2$, $L-N_2$, and $L+N_2$) of Group Two (see Appendix 18).

Each of the stimulus sets was balanced for speakers' L1, gender, and speech conditions. Half the utterances (12 true, 12 false) were spoken by the native English speakers, and half were spoken by the native speakers of Cantonese. Of the sentences spoken by each group of speakers, twelve (6 true, 6 false) were produced by female speakers and the others were produced by male speakers. Moreover, six utterances (3 true, 3 false) were presented in each of the four speech conditions (i.e., Q-N, Q+N, L-N, and L+N). For each speech condition, one of the triplets (either 3 true or 3 false) was spoken by 3 female speakers, and the other by 3 male speakers.

An overall balanced distribution of speech conditions for the true and false sentences produced by each of the speakers in the four stimulus sets was designed (see Appendix 18). As already mentioned, a pair of sentence tokens (1 true, 1 false) randomly selected from a particular speaker in one set was identical to the pair in the other set of the same group. However, the sentence tokens were completely different from those of the two pairs that were selected from the two sets of the other group. Furthermore, within the same group, the four sentences in the sets were all presented in the four different speech conditions. If any sentence (either true or false) was "Clean" speech in the first set, the corresponding true or false sentence produced by the same speaker in the other set should then be "Noisy" speech, or vice versa. In addition, each of these four sentences (either true or false) was selected from the four lists of stimuli in the two groups so that the 4 true and 4 false utterances by the same speaker could be presented in the four speech conditions (i.e., across the two groups).

To summarize, the stimuli fell into eight categories (two groups of speakers and four speech conditions): (1) native Cantonese Quiet speech without noise (NC: Q-N), (2) native Cantonese Quiet speech with noise (NC: Q+N), (3) native Cantonese Lombard speech without noise (NC: L-N), (4) native Cantonese Lombard speech with noise (NC: L+N), (5) native English Quiet speech without noise (NE: Q-N), (6) native English Quiet speech with noise (NE: Q+N), (7) native English Lombard speech without noise (NE: L-N), and (8) native English Lombard speech with noise (NE: L+N). In addition, the eight listeners in groups of two would be randomly assigned to listen to one of the four sets of stimuli (i.e., Set 1A, Set 1B, Set 2A, or Set 2B).

3.1.3. Listening Procedures

Individual listening sessions were held in a sound-treated room. Stimuli were presented at a comfortable listening level through Technics RP-HT400 headphones via a Power Macintosh computer using a custom response-collection program (Munro, 1997). The entire stimulus set was played two times with the tokens being randomized for each listener. During the first pass for the intelligibility tasks, after hearing an item, the listeners were expected to write out the statement in standard orthography on booklets with numbered spaces for transcriptions of each of the 48 utterances (Appendix 19). In the event that the listeners understood only part of the sentence, they were told to write out as many words as possible. If any word or the whole sentence was unintelligible, they should indicate this by drawing a blank line. Immediately after each orthographic transcription, the participants had to verify the truth value of the statement by selecting one of three screen buttons that could be pressed using the mouse ("True", "False", or "Unknown"). The latter was available to prevent the participants from consistently choosing "False" when an utterance was unintelligible, as in Munro (1998).

During the second pass, the listeners were asked to rate each utterance on a scale of 1 to 9 for both comprehensibility and foreign (i.e., non-English) accentedness by pressing numbers on a computer keyboard. A scale of this size had previously been used in other experiments (Derwing & Munro, 1997; Munro & Derwing, 1994, 1995a, 1995b, 1998, 1999), and was found to be effective for eliciting judgements of nonnative speech (Munro & Derwing, 1995a). Four listeners rated accentedness before comprehensibility, and the other four performed the ratings in the opposite order. For the accentedness ratings, the listeners were instructed to decide how strong each speaker's foreign accent was. A rating of 1 was used for speakers who had no foreign accent at all, while 9 was used for speakers who had a very strong foreign accent. For the comprehensibility ratings, the listeners were instructed to judge how difficult each speaker was to understand. A rating of 1 was used if the speaker was not difficult to understand, and 9 was to be used if the speaker was very difficult to understand. The labels, "no accent" and "very strong accent" for the accentedness ratings, "not difficult to understand" and "very difficult to understand" for the comprehensibility ratings, were displayed at the ends of the rating scale on the screen as reminders (Munro & Derwing, 1998). Numbers in between were used for intermediate degrees of accent and comprehensibility. Prior to performing the two rating tasks, the listeners were advised to try to use the full scale in their ratings by assigning some scores at both extremes (i.e., some scores of 1 and 9). This should not be difficult because they had already listened to the complete stimulus set in the transcription and verification tasks. They should at least have some idea of the range of accentedness or comprehensibility exhibited by the speakers.

Before completing the actual experiment, the listeners completed a practice session during which they heard a random presentation of eight statements not used in the actual experiment (see Appendix 20). Four were true sentences, and four were false sentences. Half of the true and half of the false items were mixed with the masking noise in the same manner as the experimental stimuli. The sentences were produced by four new speakers, two (1 female, 1 male) of whom were native speakers of Canadian English and two (1 female, 1 male) of whom spoke with a Cantonese accent. Each speaker was represented by one true statement and one false statement. This practice set served to familiarize the listeners with the masking noise in the speech samples, and with the experimental procedure. Immediately afterwards, they completed the listening tasks. A 5-minute break was given between the two passes, and the total time required averaged about 45 minutes.

3.2. Results

3.2.1. Intelligibility

The assessment of intelligibility of the speech samples was based on the listeners' scores on two different tasks: transcription and verification. For the transcription task, two scoring procedures were employed. In an exact word match (Derwing & Munro, 1997; Munro & Derwing, 1995a, 1995b; Munro, 1998), a correct transcription had to correspond exactly to the actual sentence. In a key word match, only correctly-identified content words, such as nouns, verbs, adjectives, and adverbs (Munro & Derwing, 1995a;

Palmer, 1981: 130; Quach, 1998), but not function words, in the sentence were counted. Content words are more important for intelligibility because they carry most of the "content " of the sentences. This procedure was undertaken to compare the results with those of the exact word match. The latter task is more rigorous because it penalizes the listeners for the transcription mistakes that may not be related to intelligibility. All the words considered content words are italicized and shown in Appendix 3. In both tasks, scores were assigned to each transcribed sentence by computing the percentage of words (or content words) in the sentence that were correctly written out. Certain minor errors, however, were ignored (e.g., the use of a singular form instead of a plural or any trivial spelling mistake). For the verification task (Munro, 1998), the scores were determined by summing the number of correct true and false responses; the choice of "Unknown" was considered incorrect.

Since it has been reported in one study (Pisoni & Dedina, 1986) that truth value (i.e., True or False) can play a role in a sentence verification task, transcription (exact word and content word matches) and verification scores by listeners across the four speech conditions (i.e., Q-N, Q+N, L-N, and L+N) for the true condition were compared with those for the false condition. Figure 3-2 and Figure 3-3 show the mean scores for the two transcription tasks, while Figure 3-4 shows the correct verification scores between the two types of sentences. The data were submitted to two-way repeated measures ANOVAs with Truth Value (2 levels) and Speech Condition (4 levels) as within-subjects factors. Results revealed that no significant differences were found between the true and false utterances across the four speech conditions in the exact word match (M = 90% for T and M = 87% for F, overall) and content word match (M = 88% for T and M = 85% for F, overall), Fs (1, 7) = 0.84 and 1.17, $ps > 0.05$. Nor was there a significant interaction between Truth Value and Speech Condition found for each of the exact word and content word matches, Fs (3, 21) = 1.24 and 1.77, $ps > 0.05$. However, a significant effect of Speech Condition was found, F (3, 21) = 15.76 for the exact word match and F (3, 21) = 20.34 for the content word match, $p < 0.0001$. A parallel test on the verification scores for the true (M = 82%) and false sentences (M = 76%) also yielded a nonsignificant effect of Truth Value, F (1, 7) = 2.04, $p > 0.05$, as well as a nonsignificant two-way interaction, F (3, 21) = 1.21, $p > 0.05$. Again, a significant effect of Speech Condition was found, F (3, 21) = 27.13, $p < 0.0001$. These results were similar to the findings in Munro and Derwing (1995b) and Munro (1998) in that no significant differences were observed between the true and false sentences in transcription and verification tasks across the different speech conditions overall. As a result, all subsequent statistical analyses in this study were carried out by pooling the results for the two types of sentences.

3.2.1.1. Transcription Scores

3.2.1.1.1 Analysis by Listeners

The mean scores (in percent) of the exact word match (EWM) and content word match (CWM) (pooled over the eight listeners) for the two groups of speakers, native English (NE) and native Cantonese (NC), under the four speech conditions (i.e., Q-N, Q+N, L-N, and L+N) are shown in Figure 3-5 and Figure 3-6, respectively. It can be readily seen that there were differences in transcription scores for the NE and NC sentences in both scorings. Overall, the transcription scores on the Cantonese-accented utterances were lower than those on the native English utterances in all the speech conditions. The exact-word matching scores for each listener were submitted to a two-way repeated measures ANOVA with Native Language of Speakers (NL) and Speech Condition (SC) as within-subjects factors. A significant effect of NL, $F\ (1,7) = 54.99$, $p < 0.01$, was observed, indicating that the native English speakers' utterances were correctly transcribed more often (95% correct) than those of the native Cantonese speakers (82% correct). It was also found that SC had a significant effect, F (3,21) = 15.06, $p < 0.0001$, indicating different performance from the listeners under the four speech conditions. However, no significant interaction between NL and SC was found, F (3,21) = 2.68, $p > 0.05$, suggesting that the listeners performed in a similar fashion in transcribing the sentences produced by the two groups of speakers in each of the four conditions.

The pattern of results in the content word match was similar to that observed in the exact word match. Of the 192 sentences produced by the native English speakers, 95% were transcribed correctly. Another two-way repeated measures ANOVA using the same factors this time on the content word match revealed that overall, significantly fewer of the Cantonese speakers' utterances (78%) were transcribed correctly, F (1, 7) = 107.32, $p < 0.0001$.[1] Again, the effect of SC was found to be significant, F (3, 21) = 20.34,

[1] It is conceivable that the use of a single factor (Speech Condition) throughout the perceptual analyses might obscure the separate effects of the Speech Production Condition (SP), Lombard and normal, and the Speech Listening Condition (SL), Noise and Quiet. To investigate this possibility, the data from the CWM were also submitted to a three-way repeated measures ANOVA with NL, SP and SL as within-subjects factors. The analysis revealed significant effects of NL, SL, NL x SL, and SP x SL, Fs (1, 7) = 107.32, 38.24, 10.77, and 13.49, $ps < 0.05$. However, the effects of SP, NL x SP, and NL x SP x SL were nonsignificant, Fs (1, 7) = 1.86, 0.056 and 0.00, $ps > 0.05$. Tests of simple main effects indicated significant differences in scores between native- and nonnative-produced utterances presented in quiet and noisy conditions, and in scores between the two speech listening conditions for sentences produced by both groups of speakers ($ps < 0.05$). Moreover, Lombard speech was correctly transcribed more often than normal speech when presented in noisy conditions ($p < 0.05$), but not in quiet conditions ($p > 0.05$). Both Lombard and normal speech received significantly higher scores when presented in quiet than in noisy conditions ($ps > 0.05$). These findings are compatible with the results obtained in the two-way repeated measures ANOVA.

$p < 0.0001$, indicating that, as in the exact word match, the listeners performed differently under the four speech conditions in transcribing the sentences. The interaction NL x SC, however, failed to reach significance, $F(3, 21) = 20.34$, $p > 0.05$.

Post hoc analyses were carried out to examine the listeners' performance under each of the four speech conditions. Overall, the sentences in the Q-N condition were correctly transcribed at a higher mean rate ($M = 96\%$ for EWM and 95% for CWM) than were the utterances in L-N ($M = 95\%$ for EWM and 94% for CWM), L+N ($M = 86\%$ for EWM and 82% for CWM), or Q+N ($M = 77\%$ for EWM and 75% for CWM). Post hoc Fisher's PLSD tests were used to determine which pairwise differences among the speech conditions were significant. For both the exact word match and the content word match, results revealed that utterances presented in Q+N and L+N conditions received significantly lower scores than those in either Q-N or L-N, $ps < 0.05$. It was also found that listeners' scores in the L+N condition were significantly higher than those in the Q+N condition, $ps < 0.05$, but that the difference between L-N and Q-N proved nonsignificant, $ps > 0.05$. These results suggest that the added masking noise degraded the intelligibility of both the Quiet (i.e., normal) and the Lombard speech. Moreover, the Lombard speech was significantly more intelligible than the Quiet speech in the noisy, but not in the noise-free, condition.

3.2.1.1.2 Analysis by Individual Speakers

The data from the two tasks were also examined from the standpoint of individual speakers to have a better understanding of between-speaker effects. The amount of degradation of the Quiet and the Lombard speech for each speaker rendered by the masking noise can be evaluated by studying individual scores in the four speech conditions. The mean transcription score for each speaker under each of the speech conditions was computed by averaging the scores of one true sentence and one false sentence, each of which was scored by two listeners. Since the results of the exact word match and the content word match showed a similar pattern, only the data for the content word match were presented here. The individual speakers' mean scores for the native Cantonese speakers are given in Figure 3-7 and those for the native English speakers are given in Figure 3-8. As in Munro's (1998) examination of Mandarin speakers' English sentences perceived in cafeteria noise, several of the Cantonese speakers received very low transcription scores. Four of them (C-f2, C-f3, C-f6, and C-m4) in the Q+N condition obtained scores of 50% or below (50%, 31%, 44% and 46%, respectively), but none of the English speakers did. The lowest transcription scores obtained by two native English speakers (E-m4 and E-m5) were both 56% under the Q+N condition.

As was the case in Munro (1998), some speakers in both groups produced sentences that were highly resistant to noise. That is, there was no decrement in intelligibility scores for sentences being presented in the clean and the noisy conditions. Four of the native English speakers (E-f1, E-f3, E-m1, E-m6) showed 100% intelligibility throughout the four speech conditions. Others received full scores either in the Q-N and Q+N conditions (E-m2) or in the L-N and L+N conditions (E-f2, E-m3, E-m4, E-m5). For the nonnative speakers, one female speaker (C-f1) obtained scores identical in the Q-N and Q+N conditions (100%) and another female speaker (C-f5) in the L-N and L+N conditions (81%). The male Cantonese speaker, C-m1, received 100% in both the Q-N and Q+N conditions and 83% in both the L-N and L+N conditions.

It was interesting to note that one Cantonese speaker received transcription scores higher in the noisy speech condition than in the clean speech condition, whereas none of the native English speakers did. The female speaker (C-f3) showed a score (67%) higher in the L+N condition than in the L-N condition (42%).

It was also noteworthy that obtaining a high transcription score in the clean speech condition was not necessarily an indication of receiving a high score in the noisy speech condition (Munro, 1998). For those Cantonese speakers of English (C-f6, C-m2, C-m3) who received a CWM score of 100% in the Q-N condition, their scores in the Q+N condition were 44%, 67%, and 83%. The speakers C-f2, C-f6, and C-m6 received a score of 100% in the L-N condition, but scores of 67%, 58%, and 90% in the L+N condition.

An informal analysis was also performed in an attempt to explore any relationship existing between Cantonese speakers' performance in the production experiment and in the intelligibility task. Speaking rates and fundamental frequency values of those speakers whose utterances were either highly susceptible (i.e., becoming unintelligible) or highly resistant (i.e., remaining intelligible) to the masking noise in the content word match were examined. Nevertheless, no systematic patterns in the two respective production parameters were found for the speakers belonging to either the "noise-susceptible" or the "noise-resistant" group. For instance, the two female Cantonese speakers (C-f3 and C-f6) whose utterances in the Q+N condition received the lowest transcription scores each produced the corresponding sentences in the Quiet speaking condition at rates which were either faster or slower than was the overall mean speaking rate for all the female Cantonese speakers. The values of the mean sentential Fo and Fo range of the same sentences in the Quiet condition for the two speakers also showed similar patterns (i.e., either higher or lower than the overall mean Fo and Fo range values). For the two Cantonese speakers (C-f1 and C-m1) who obtained a score of 100% in the Q-N and Q+N conditions, no regular patterns, as in the case for the "noise-susceptible" group, were found in their speaking rates and fundamental frequency values of the relevant sentences produced in the Quiet condition.

3.2.1.2. Verification Scores

Verification scores (i.e., the number of times that each subject assigned correct truth value to the sentences) in each of the four speech conditions (i.e., Q-N, Q+N, L-N, L+N) were tabulated for each listener. As summarized in Figure 3-9 which gives the mean verification scores (in percent) for the two groups of speakers (Cantonese and English), it can be seen that there were important differences in verification scores for the NC and NE utterances. Overall, sentences spoken by the native Cantonese speakers were correctly verified much less frequently than were the sentences spoken by the native English speakers in each of the four speech conditions. Data from the eight listeners were submitted to a repeated measures ANOVA with Native Language of Speakers (NL) and Speech Condition (SC) as within-subjects factors. The statistical analysis yielded significant main effects of NL, F (1, 7) = 68.03, $p <$ 0.0001, and SC, F (3, 21) = 27.14, $p < 0.0001$, as well as a significant NL x SC interaction, F (3, 21) = 3.24, $p < 0.05$. Two separate one-way repeated measures ANOVAs with SC as the within-subjects factor were carried out to examine the listeners' performance under each of the four speech conditions on verifying utterances produced by the native Cantonese and the native English speakers. For the sentences produced by the native Cantonese speakers, results of the one-way repeated measures ANOVA showed a significant difference in SC, F (3, 21) = 14.86, $p < 0.0001$. Post hoc Fisher's PLSD tests revealed that sentences presented in both Q-N (M = 81%) and L-N (M = 81%) were correctly verified more often than those presented in either Q+N (M = 50%) or L+N (M = 52%), $ps < 0.05$. However, all other pairwise comparisons (L-N vs. Q-N and L+N vs. Q+N) failed to reach significance, $ps > 0.05$. The pattern of the results for the verification scores suggest that both the Lombard and the Quiet speech produced by the Cantonese speakers of English are more difficult to verify in noisy than in quiet conditions. However, the Cantonese speakers' Lombard speech is not significantly different from their normal speech under quiet and noisy conditions when verified by the native English listeners.

For the native English productions, the mean scores were submitted to another one-way repeated measures ANOVA with Speech Condition as a within-subjects factor. As was the case with the nonnative sentences, a significant effect of SC was observed, $F(3, 21) = 7.54$, $p < 0.01$. Results of Fisher's PLSD tests showed that utterances presented in Q+N received significantly lower scores ($M = 79\%$) than did those in Q-N ($M = 100\%$), L-N ($M = 96\%$), or L+N ($M = 92\%$), $ps < 0.05$. Nevertheless, there were no significant differences among the three speech conditions, Q-N, L-N, or L+N, $ps > 0.05$. These show that the noisy Quiet speech was correctly verified less often than the clean Quiet speech, the clean and noisy Lombard speech. Moreover, it is interesting to note that despite a nonsignificant difference in scores between L-N and L+N, the quiet Lombard speech tended to be correctly verified at a somewhat higher rate than the noisy Lombard speech.

On the basis of all the above analyses, the interaction was due to the fact that Cantonese-accented Lombard speech simply does not seem to have the properties needed to overcome the effects of noise, whereas native speech does.

3.2.2. Rating Tasks

3.2.2.1. Accent Ratings

With reference to the procedure used by Derwing and Munro (1997) and Munro and Derwing (1999) for determining how well the listeners agree with one another in rating accentedness, the overall inter-rater reliability was assessed using the method suggested by Hatch and Lazaraton (1991). This approach involves computing mean inter-rater correlation coefficients after conversion to Fisher Z scores. Since two listeners were randomly assigned to rate one of the four sets of stimulus lists (i.e., Sets 1A, 1B, 2A, or 2B), the coefficients were computed for each of the paired listeners rating the correlated stimulus sentences in the same set. Results revealed Pearson (r) values of 0.77 for Set 1A, 0.83 for Set 1B, 0.83 for Set 2A, and 0.81 for Set 2B. According to Haimson and Elfenbein (1985), a correlation between ±0.80 and ±0.90 is generally considered high. As a result, the above Pearson (r) values suggest a high degree of inter-rater agreement among the listeners in the accent ratings.

An examination of the listeners' accent ratings revealed that overall, the native English (NE) speakers' productions always received much lower ratings (M = 1.4) than did those of the native Cantonese (NC) speakers (M = 6.5) on the nine-point scale. Figure 3-

10 gives the mean accent ratings assigned by the eight listeners to both groups of speakers in each of the four speech conditions (Q-N, Q+N, L-N, L+N). A two-way repeated measures ANOVA on the listeners' data, with Native Language of Speakers (NL) and Speech Condition (SC) as within-subjects factors, indicated a significant effect of NL only, F (1, 7) = 208.38, $p < 0.0001$. The effect of SC and the interaction of the two factors were both nonsignificant, F (3, 21) = 2.05, and F (3, 21) = 0.18, $ps > 0.05$, respectively. These data show that the Cantonese speakers, as expected, are rated as producing significantly more accented speech than their English counterparts in all the speech conditions, and that the different speech conditions have no significant effect on the accent ratings of speech produced by the two groups of the speakers.

3.2.2.2. Comprehensibility Ratings

As was the case in the accent ratings, the inter-rater reliability for the comprehensibility ratings was determined using the method recommended by Hatch and Lazaraton (1991). The inter-rater correlation coefficients for the pairs of listeners in the four sets were 0.72 (Set 1A), 0.76 (Set 1B), 0.73 (Set 2A), and 0.75 (Set 2B), respectively. With reference to Haimson and Elfenbein (1985), these values overall indicate a moderately high degree of agreement, slightly lower than that observed for the accent ratings.

Comprehensibility ratings were tabulated for each of the 384 stimulus sentences. Figure 3-11 shows the mean ratings (pooled over the eight listeners) for the sentences produced by the two groups of speakers under each of the four speech conditions. It can be readily seen that utterances produced by the Cantonese speakers consistently received higher ratings (M = 4.7) than did the English speakers' productions (M = 1.8) overall. It was found that the comprehensibility ratings for NC (M = 5.9) and for NE (M = 2.7) in the Q+N condition were higher than those for NC (M = 3.6) and for NE (M = 1.2) in the speech condition of Q-N. Furthermore, the assigned ratings for NC (M = 5.7) and for NE (M = 2.3) in the L+N condition were higher than those for NC (M = 3.6) and for NE (M = 1.2) in the L-N condition. A parallel ANOVA on the comprehensibility ratings revealed significant effects of NL, F (1, 7) = 53.41, $p < 0.01$, and of SC, F (3, 21) = 22.77, $p < 0.0001$. No significant two-way interaction, however, was found, F (3, 21) = 1.62, $p > 0.05$. Post hoc Fisher's PLSD tests were used to determine which pairwise differences between the four speech conditions were significant. It was found that utterances presented in both Q+N and L+N conditions received significantly higher scores (i.e., less comprehensible) than those presented in the conditions of either Q-N or L-N, *ps* < 0.05. Nonetheless, all other differences (Q-N versus L-N and Q+N versus L+N) failed to reach significance, *ps* > 0.05. These results suggest that in all the speech conditions, the English sentences produced by the native Cantonese speakers were considerably more difficult to understand than were those spoken by the native English speakers. In addition, both the Lombard and Quiet speech produced by either the Cantonese or English speakers were less comprehensible in the noisy than in the quiet conditions. However, the Lombard speech was not more comprehensible than was the Quiet speech when perceived by the native English listeners in both the noisy and the noise-free conditions.

3.2.3. Cross-Task Comparisons

Interrelationships among the different kinds of tasks were explored using correlational analyses. In this experiment, a pair of listeners was randomly grouped and assigned to listen to correlated tokens in one of the four sets of stimuli. Each of the eight listeners had to transcribe, verify, assign accent and perceived comprehensibility ratings to the 48 sentences. A Pearson correlation coefficient (r) between two different tasks for all the speakers was computed. Because the results of eight correlation tests were performed, a conservative alpha level of 0.01 was adopted. Since parallel results were obtained in the verification and transcription tasks in at least one recent study (Munro, 1998), correlation between these two tasks would not be examined here.

For each of the eight native English listeners, the correlation between the exact word match and content word match was highly significant (p <0.0001), with a range of 0.87 to 0.97. The results indicated very small differences between the two types of orthographic transcription tasks in measuring intelligibility. Parallel comparisons were also made between the mean accent and perceived comprehensibility ratings, and most listeners showed a significant Pearson correlation ($p < 0.0001$). However, for individual listeners the strength of the correlation varied considerably ranging from 0.29 to 0.91. For two of the eight listeners, the correlation failed to reach significance at the 0.01 level. These findings were highly similar to those of Munro and Derwing (1995a, 1995b) in that accentedness and comprehensibility are only partially-correlated dimensions of L2 speech. Table 3-1 shows all the correlation coefficients and their respective significant levels obtained from the eight listeners for the aforementioned two comparisons (i.e., exact word match and content word match, assigned accent and comprehensibility ratings).

3.2.4. Transcription Errors

Following Munro (1998), the nature of the errors in the transcription data was examined. As was the case in Munro (1998), most of the errors that the listeners made in this study broadly fell into two types: (1) missing word(s), or even a whole sentence, and (2) mistranscribed word(s) or sentence(s). Of the 384 sentences transcribed by the 8 native English listeners, 10 (about 2.6%) received scores of 0. Table 3-2 gives the numbers of these completely wrong transcriptions according to the stimulus conditions. Of the 10 completely incorrect transcriptions, nine of them were produced by the native Cantonese speakers, and only one was produced by a native English speaker (in the condition of Q+N). Of the 9 nonnative productions, eight of them were presented in the noisy speech conditions (4 in the Q+N condition, and 4 in the L+N condition), and the other one was presented in the condition of L-N. The data were submitted to a chi-square test to determine the relationship between the number of the zero-scored transcriptions and the stimulus conditions. Results revealed that the distribution of the number of the completely wrong transcription across the stimulus condition was nonrandom, $\chi^2(7) = 17.20$, $p < 0.05$.

According to Munro (1998), mistranscription occurred when the listeners were unaware of misinterpreting a word or a series of words produced by the speaker. It was likely that the listeners genuinely believed they understood something that the speaker had said, because they pressed either the "True" or "False" button, but not the "Unknown" one. They selected the latter button only when they did not understand the sentence or were uncertain of the truth value of the sentence. Out of the 384 sentences, a total of 25 (7%) were identified as mistranscribed utterances. In some instances, the mistranscription happened to lead to a correct verification. For example, a listener mistranscribed the sentence "Young children can be very noisy" (a true sentence) as "Young student can be very noisy", and she verified it as true. Another two listeners transcribed the target sentence "You can tell time with a kettle" (a false statement) as "You can tell time with a castle", and verified it as false. In contrast, the wrong truth value was assigned in other instances. For example, the true sentence "Some people keep dogs as pets" was written and verified as a false sentence "Some people keep salt as pets". The false sentence "It's good to have stones for breakfast" was interpreted as "It's good to have scones for breakfast" and was verified as true. Table 3-2 also shows the frequencies of mistranscriptions according to the stimulus condition. A chi-square test indicated that the distribution of the numbers of mistranscribed sentences across the stimulus conditions was non-random, $\chi^2(7) = 15.00$, $p < 0.05$. Among the mistranscribed sentences, 19 sentences (i.e., 76%) were produced by the native Cantonese speakers, whereas the remaining 6 sentences were spoken by the native speakers of English. It was interesting to note that for the Cantonese-accented mistranscribed sentences, the number of errors was not higher in the noisy speech conditions (16% in the NC: Q+N, 16% in the NC: L+N) than in the clean speech conditions overall (12% in the NC: Q-N, 32% in the NC: L-N). This pattern of results was comparable to Munro's (1998) findings in that the number of Mandarin-accented mistranscription errors was higher in noise-free (44%) than in noisy conditions (34%). However, the reverse was true for the native productions here. As in Munro's findings (22% in noisy condition and 0% in the noise-free condition), the mistranscription errors for the native utterances in this experiment were higher in the noisy speech conditions (16% in the Q+N condition, and 8% in the L+N condition) than in the clean speech conditions (0% in both the Q-N and L-N conditions). It should be mentioned that the mistranscription errors were almost twice as common in the Cantonese speakers' clean speech (44% overall) than in the English speakers' noisy speech (24%). This finding implies qualitatively different effects of accent and masking noise on perception of English sentences. In agreement with Munro (1998), foreign accent was more likely to cause mistranscriptions than was masking noise of the kind and level used in this experiment.

3.3. Discussion

In this perception experiment, it was found that as expected, the Cantonese speakers' utterances were less intelligible and comprehensible, but more accented, than were the native English productions in each of the four speech conditions. In addition, the findings revealed that overall, the noisy sentences showed poorer scores/ ratings than did the clean speech samples, and that in some cases, the Lombard speech was better perceived than was the normal speech in the noisy conditions. However, no significant difference was found between normal and Lombard speech, when presented in clean conditions, for all three parameters.

First of all, as in previous L2 speech studies involving nonnative speakers of English (Anderson-Hsieh & Koehler, 1988; Derwing & Munro, 1997; Derwing, Munro & Wiebe, 1998; Munro, 1995, 1998; Munro & Derwing, 1994, 1995a, 1995b, 1998, 1999), the English spoken by the native Hong Kong Cantonese speakers here was perceived to be significantly different from that of the native English speakers in terms of the three dimensions - intelligibility, foreign-accentedness and comprehensibility. In each of the four speech conditions (i.e., Q-N, Q+N, L-N, and L+N), the Cantonese speakers' sentences were less correctly transcribed than those of the native productions according to both of the two scoring methods. It was found that overall, the transcription scores for the Cantonese speakers were 13% (exact word match) and 17% (content word match) less than those for the English speakers. More specifically, the sentences that received, in the content word match, the lowest mean transcription scores (50% or below, all in the condition of Q+N) were all produced by four of the Cantonese speakers, but not by any member of the native English group, the two lowest scores of which were both 56% (again in the Q+N condition). For the verification task, it is not surprising to find that the sentences produced by the Cantonese speakers were correctly verified less frequently than the speech samples produced by the English speakers in each of the four speech conditions. As was the case in the transcription task, the sentences spoken by the Cantonese speakers received the lowest verification score (50% in the Q+N condition), compared to the lowest one for the English speakers (79% in the same speech condition).

For the other two dimensions, the eight English native-speaker listeners agreed quite strongly among themselves in assigning ratings of accentedness and comprehensibility for the two groups of speakers. The speech of the Cantonese speakers was always rated as much more accented than the English speakers' utterances throughout the four speech conditions. Furthermore, the listeners found the nonnative sentences more difficult to understand than the native productions in each of the speech conditions. These findings demonstrate that the native English-speaking listeners can reliably differentiate foreign-accented English from the native English on the bases of intelligibility, accentedness and comprehensibility not only in clean conditions, as in other studies, but in noisy conditions as well.

Moreover, the listeners responded in an interesting way to the sentence-length utterances spoken in the two different speech styles (Lombard and normal) that were presented with or without the masking noise. First and foremost, it is apparent that in some cases, sentences presented without noise were better perceived than those presented with noise. For both the exact-word and content-word matching, sentences presented in the conditions of Q-N and L-N were more correctly transcribed than were sentences presented in the Q+N and L+N conditions. As expected, the additive noise does degrade the intelligibility of both normal and Lombard speech, as far as this kind of task (i.e., standard orthographic transcription) is concerned. A similar pattern of results was also observed for the comprehensibility-rating task; the clean stimulus sentences are easier to understand than are the sentences presented with noise. It is likely that the masking noise exerts a detrimental effect on the listeners in understanding the English sentences, regardless of the native language of the speakers.

For the verification task, the expected pattern that the clean speech should be correctly verified more often than the noisy speech was again observed in the scores for the sentences spoken by the Cantonese speakers, but not for those spoken by the native speakers of English, however. For the latter, the normal speech was correctly verified more frequently in the noise-free condition than in a noisy condition. In contrast, the noisy native English Lombard speech was correctly verified at a mean score (92%) almost as high as that of the clean Lombard sentences (96%).

It is surprising that normal sentences presented without noise were perceived to be as good as Lombard sentences presented in the noise-free condition. This pattern of results can be readily observed in all the tasks employed in this experiment - transcription, verification, and the rating tasks for foreign-accentedness and comprehensibility. Absence of perceptual difference can readily be observed in similar scores or ratings of the speech presented in the conditions of Q-N and L-N. This may be due to the fact that in this experiment, the speaking rates and Fo values manifested in Lombard sentences do not have a significantly greater effect on intelligibility, accent and comprehensibility judgements than those observed in normal sentences, when both types of speech are presented in a noise-free environment.

Nevertheless, sentences presented in the speech condition of L+N were better perceived than those presented in the Q+N condition, though for some tasks only. For both the exact word match and content word match, the findings indicated that the noisy Quiet speech was correctly transcribed less frequently than were the noisy Lombard sentences. This shows that the Lombard speech is more intelligible than the normal speech in a background of masking noise. Another piece of supporting evidence comes from the fact that for each of the two groups of speakers, the sentences that received the lowest transcription scores in the content word match were all presented in the Q+N condition, but not in the L+N condition.

For the verification task, such a pattern was found for the native English speakers, but not for the Cantonese speakers of English. The mean verification score of the nonnative sentences in the condition of L+N (52%) was only a bit higher than that of the stimuli presented in the Q+N (50%) condition. In contrast, the native productions presented in the L+N condition obtained a significantly higher mean verification score, an increase of almost 17%, than did those presented in the Q+N condition. The findings demonstrate that Lombard speech of the native English speakers is more robust for verification in noise than that of the Cantonese speakers of English.

As far as the comprehensibility-rating task is concerned, as in the noise-free condition, the Lombard speech was not rated as easier to understand than the Quiet speech when presented in the noisy condition. Interestingly, the accent ratings assigned by the listeners between the two speech styles (i.e., normal or Lombard) presented under noisy speech condition were similar. It is then clear that for each of the two groups of speakers, the assigned accent ratings remain more or less the same throughout the four different speech conditions (i.e., Q-N, L-N, Q+N, and L+N). This indicates that Lombard speech does not differ from the normal speech in accentedness rating, and that the masking noise used in this study has no significant effect on causing the speech, especially those produced by Cantonese speakers of English, perceived to be more accented.

Nonetheless, the addition of the masking noise seems to have a more detrimental effect on scores and ratings for the sentences produced by the Cantonese speakers than for those spoken by the native speakers of English. In the content word match, for example, the mean percent correct transcription score for the Cantonese speakers' sentences had a drop of about 25% from the Q-N to Q+N conditions. From the conditions of L-N to L+N, there was a decrease of almost 18% for the foreign-accented speech. The corresponding figures for the native-produced sentences were only about 14% (Q-N to Q+N) and 7% (L-N to L+N), respectively. Similar patterns are also observed for the data on the exact word match, verification, as well as the comprehensibility-rating task. These replicate part of Munro's (1998) findings in that the effect of masking noise on the intelligibility of foreign-accented English, produced by Mandarin speakers in his experiment, is greater than that on the native productions.

The above results provide evidence that nonnative English seems to be more difficult to understand, especially when the speech samples are presented in noisy condition. However, in agreement with Munro (1998), the findings presented here do not confirm the fact that interspeaker variability is only minimal among nonnative English speakers, as reported by Lane (1963). First of all, it is noted that in the conditions of Q+N and L+N, the sentence productions of some speakers of both groups were quite resistant to the masking noise; whereas the speech of other speakers was severely affected by the background noise. It was even worse for the Cantonese speakers' sentences presented in the Q+N condition. In fact, as shown in Figure 3-7, it is apparent that for the content word match, for instance, a large degree of variability in scores for the Cantonese-accented English can be readily observed throughout each of the four speech conditions. Moreover, in agreement with Munro's (1998) findings, the scores obtained by the nonnative English speakers in clean conditions may not necessarily be an accurate indication of performance in the noisy condition. It can be readily seen that in the content word match, there were some Cantonese speakers who received 100% in either Q-N or L-N conditions, but obtained scores from as low as 44% to 90% in the corresponding noisy conditions (i.e., Q+N or L+N). Consequently, as suggested by Munro (1998), the effect of masking noise on the perception of foreign-accented speech cannot be generalized in such a simple way as Lane had suggested about four decades ago. The discrepancy in the findings of the degree of interspeaker variability between Lane (1963) and the current study, as well as Munro (1998), may be due to the type of noise and different number and gender of nonnative speakers employed in the respective experiments. Lane used only 3 male nonnative speakers of three different L1s (Japanese, Punjabi, and Serbian), whose utterances were presented in white noise. In contrast, both Munro (1998) and the present study employed over 10 native speakers of Chinese (Mandarin and Cantonese, respectively) with balance for gender, whose noisy sentences were all presented in cafeteria noise.

As already mentioned, two major types of errors from the listeners' transcription data have been identified - completely incorrect transcription or mistranscription. It is interesting to note that how often the listeners made such errors seems to be dependent on the different stimulus conditions. For the type of error where the listeners obtained scores of zero in transcribing the English sentences, it occurred most often in noisy speech conditions (i.e., Q+N and L+N), especially when the sentences were spoken by the native Cantonese speakers. Only one incidence of completely unintelligible transcription occurred in a sentence produced by an English speaker presented in the condition of Q+N. This further supports the claim that nonnative English productions appear to be less intelligible than the speech produced by the native English speakers in adverse listening condition.

In the second type of error, the listeners happened to be unaware of the fact that they misunderstood what the speakers had actually said, and thus they made their choice accordingly by selecting either the "True" or "False" button, but not the "Unknown" one. As was the case for the completely wrong transcription, the incidence of mistranscriptions happened with the sentences produced by the Cantonese speakers much more frequently than those spoken by the English speakers (76% versus 24%, overall). For the native productions, the listeners mistranscribed the sentences presented in the noisy conditions (i.e., Q+N and L+N). However, for the speech produced by the Cantonese speakers, it is found that the incidence of mistranscriptions occurred not only in the noisy conditions, but in the clean speech conditions (i.e., Q-N and L-N); the latter conditions even had higher error rates for mistranscriptions. It is highly likely that under the noisy speech conditions, as suggested by Munro (1998), the listeners were likely to miss the foreign-accented sentence completely, or to press the "Unknown" button instead of the "True" or "False", rather than mistranscribe words. This is evident from the fact that the frequencies for completely wrong transcriptions were highest (40%) in both the stimulus conditions of NC: Q+N and NC: L+N.

Last but not least, it should be mentioned that overall, the Cantonese speakers' sentences presented in clean conditions caused mistranscriptions almost two times more than the native productions presented in noisy conditions do. In agreement with Munro (1998), this piece of evidence reported here supports the claim made by Lane (1963) that foreign accent is considered to be a response-dependent operation relying on listener's response during undistorted speech transmission. In Lane's view, foreign accent is different from the other kind of speech distortion caused by masking of speech signal, for example. The latter is known as response-independent operation in which the speech distorting operation is not determined by listener's response.

3.4. Conclusion

The effects of masking noise on nonnative conversational and Lombard speech have been examined in this perception experiment. Simple English true and false sentences produced in quiet and in noisy environments by the native speakers of Cantonese and English were digitally mixed with the cafeteria-like masking noise at a constant S/N ratio and presented with those noise-free sentences to a new group of native English-speaking listeners. The clean and noisy sentences, all equated for overall RMS amplitude, were assessed in terms of intelligibility, foreign-accent and comprehensibility.

In this study, nonnative English was found to be less well perceived than native–produced English. It was found that the speech of the Cantonese speakers was less intelligible than the native productions in each of the four speech conditions. Furthermore, the sentences produced by the Cantonese speakers were found to be more difficult to understand and more accented than were the sentences produced by the native speakers of English in each of the different speech conditions.

Moreover, speech samples presented without noise are better perceived than are those masked with noise in general. Sentences spoken by the two groups of speakers presented in clean conditions were frequently found to be more intelligible and comprehensible than those presented in the noisy conditions. However, the effects of the masking noise on degrading perception are far more profound for the nonnative speech than for the sentences produced by the native English speakers.

It is also observed that Lombard and normal speech interact differently in quiet and in noisy conditions. When presented in a noise-free condition, as far as all the perception tasks are concerned, Lombard and normal conversational speech seem to be as intelligible and comprehensible as each other. However, as already mentioned, Lombard speech is more intelligible than is normal speech in noisy environment in some, but not all, tasks only (e.g., transcription task). Interestingly, it has been discovered in this research that neither the speech style nor the masking noise has any significant effect on influencing the accentedness ratings assigned by the English native-speaker listeners.

The findings reported in this experiment support one of the claims, but not the other, that were made by Lane (1963) about perception of nonnative English in noise. The transcription data from the speech of the Cantonese speakers revealed considerable interspeaker variability. With reference to the mistranscription data, the distinction between the two kinds of speech distorting operations, response-dependent and response–independent (Lane, 1963), has been confirmed in this perception experiment.

TABLE 3-1

The Pearson correlation coefficients (r) for all the speakers between (A) exact-word and content-word matching scores, and (B) accent and comprehensibility ratings obtained from the eight native English listeners.

Stimulus List	Pearson r for (A)	Pearson r for (B)
Set 1A	0.92 ***	0.52 ***
	0.95 ***	0.91 ***
Set 1B	0.95 ***	0.64 ***
	0.93 ***	0.91 ***
Set 2A	0.97 ***	0.41 **
	0.91 ***	0.67 ***
Set 2B	0.96 ***	0.35 *
	0.87 ***	0.29 *

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

TABLE 3-2

Frequencies of 0% transcriptions and mistranscribed sentences.

Stimulus Condition	Completely Wrong Transcriptions (N = 48 per cell)	Mistranscribed Utterances (N = 48 per cell)
NC: Q-N	0 (0%)	3 (12%)
NC: Q+N	4 (40%)	4 (16%)
NC: L-N	1 (10%)	8 (32%)
NC: L+N	4 (40%)	4 (16%)
NE: Q-N	0 (0%)	0 (0%)
NE: Q+N	1 (10%)	4 (16%)
NE: L-N	0 (0%)	0 (0%)
NE: L+N	0 (0%)	2 (8%)
Total	10	25

FIGURE 3-1

Illustrations of waveforms of sentence no. 1 (RMS amplitude equalized at 50 dB) (top panel), the masking noise (middle panel), and the sentence mixed with the noise (bottom panel) (RMS amplitude equalized at 50 dB).

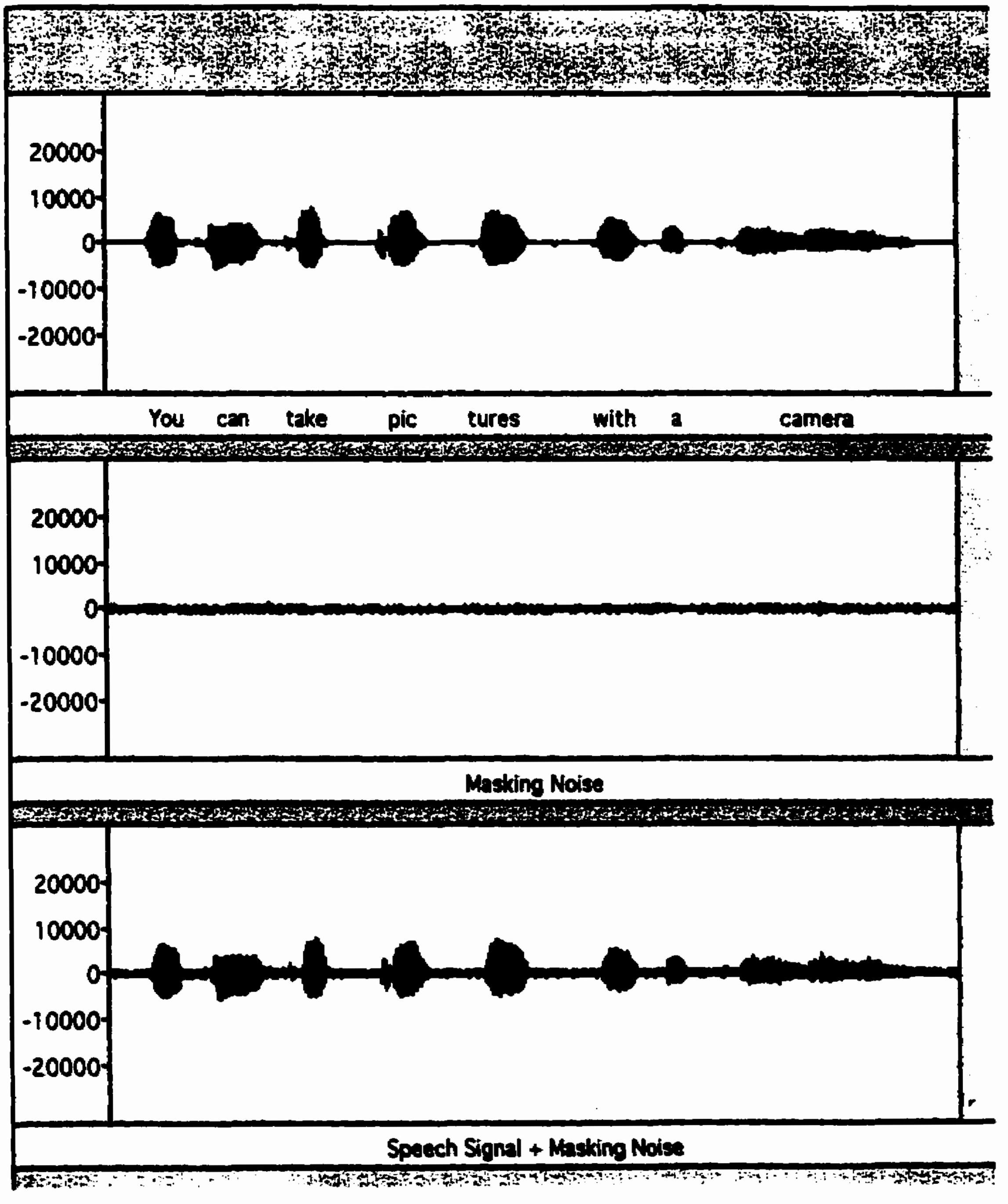


FIGURE 3-2

Percent correct transcription (exact word match) scores for the True (circles) and False (squares) utterances presented under the four speech conditions.

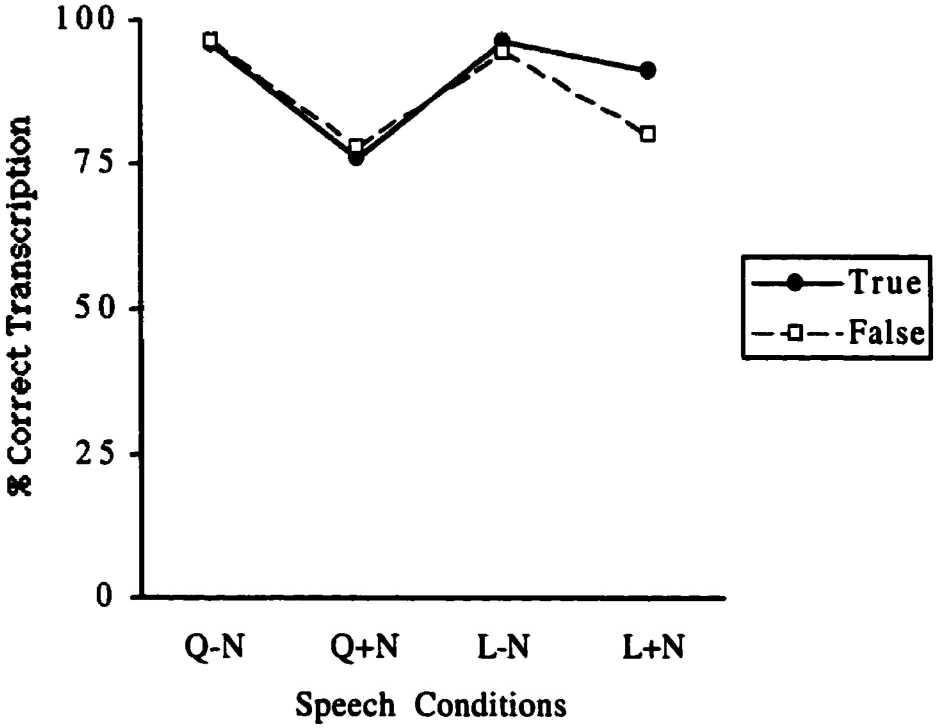


FIGURE 3-3

Percent correct transcription (content word match) scores for the True (circles) and False (squares) utterances presented under the four speech conditions.

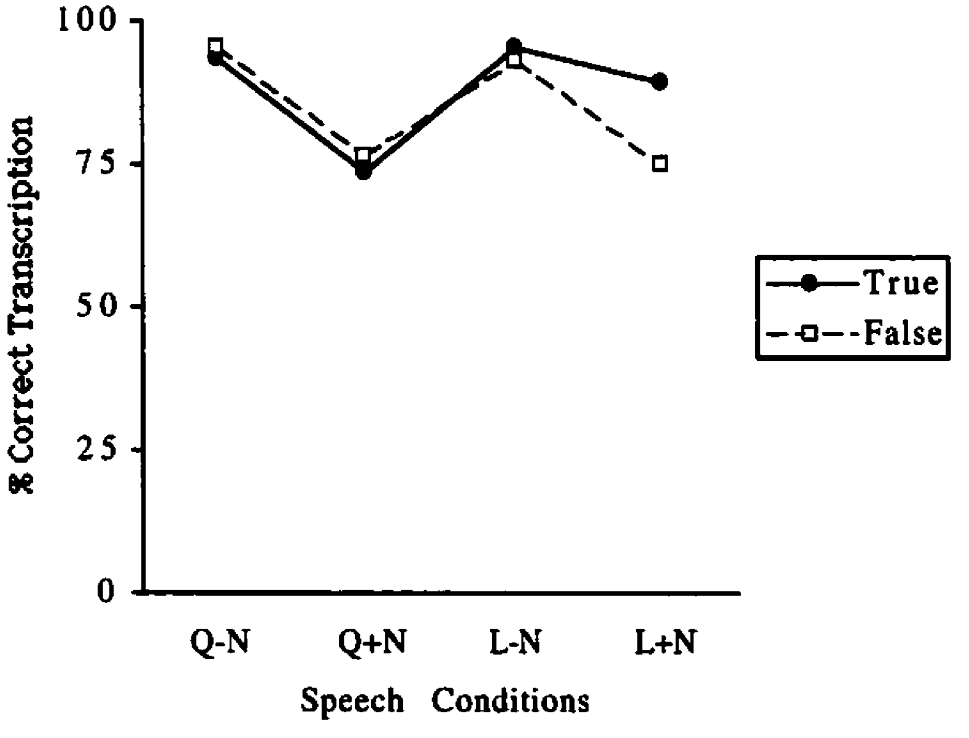


FIGURE 3-4

Percent correct verification scores for the True (circles) and False (squares) utterances presented under the four speech conditions.

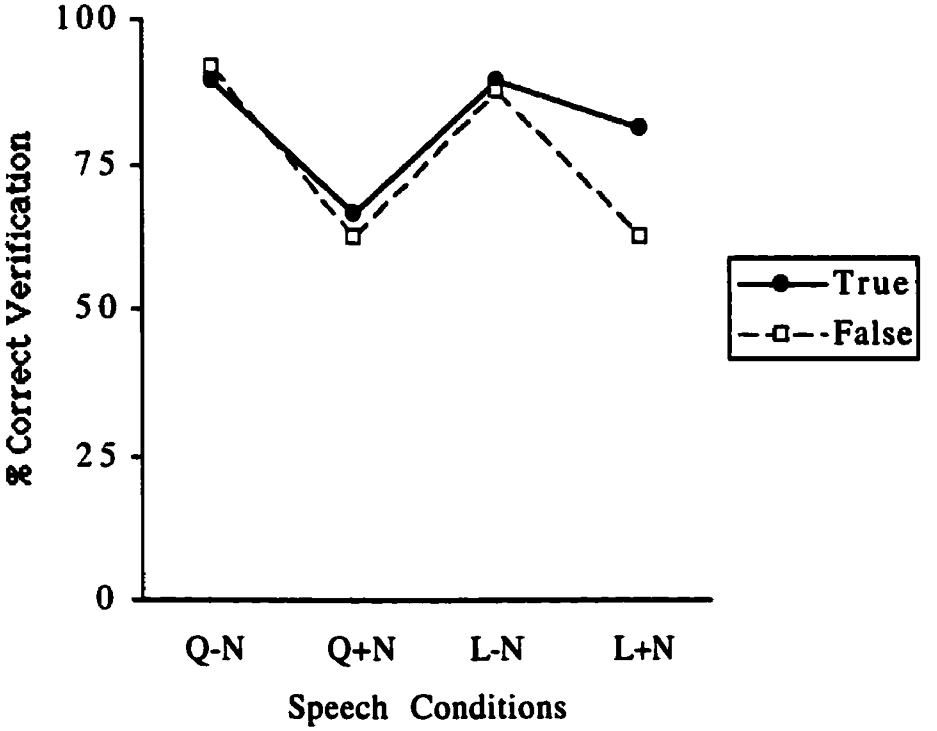


FIGURE 3-5

Percent correct transcription (exact word match) scores for the native English and native Cantonese utterances presented under the four speech conditions.

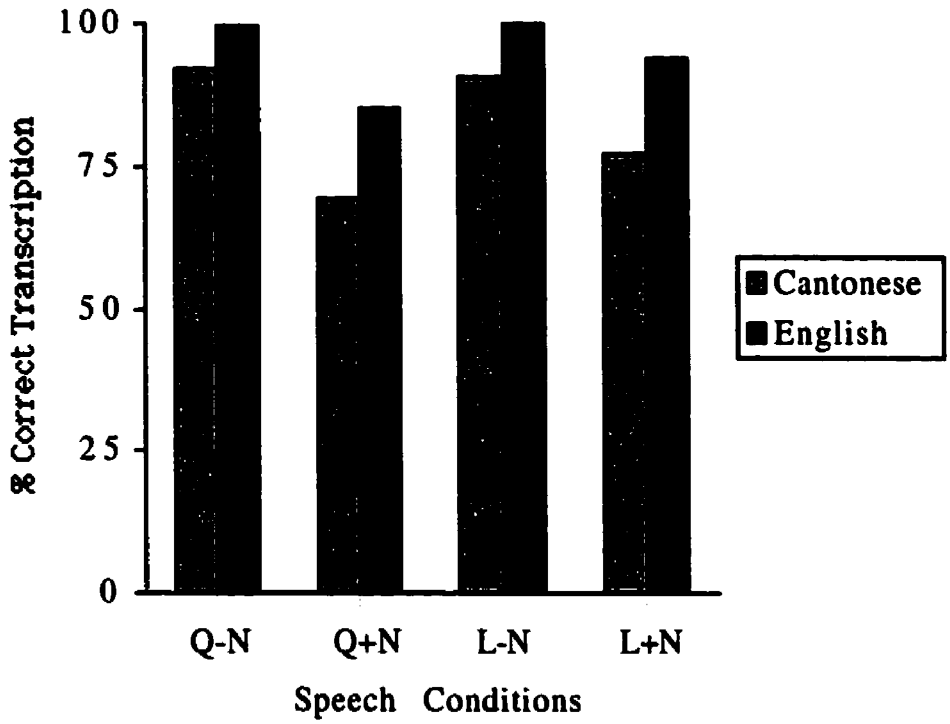


FIGURE 3-6

Percent correct transcription (content word match) scores for the native English and native Cantonese utterances presented under the four speech conditions.

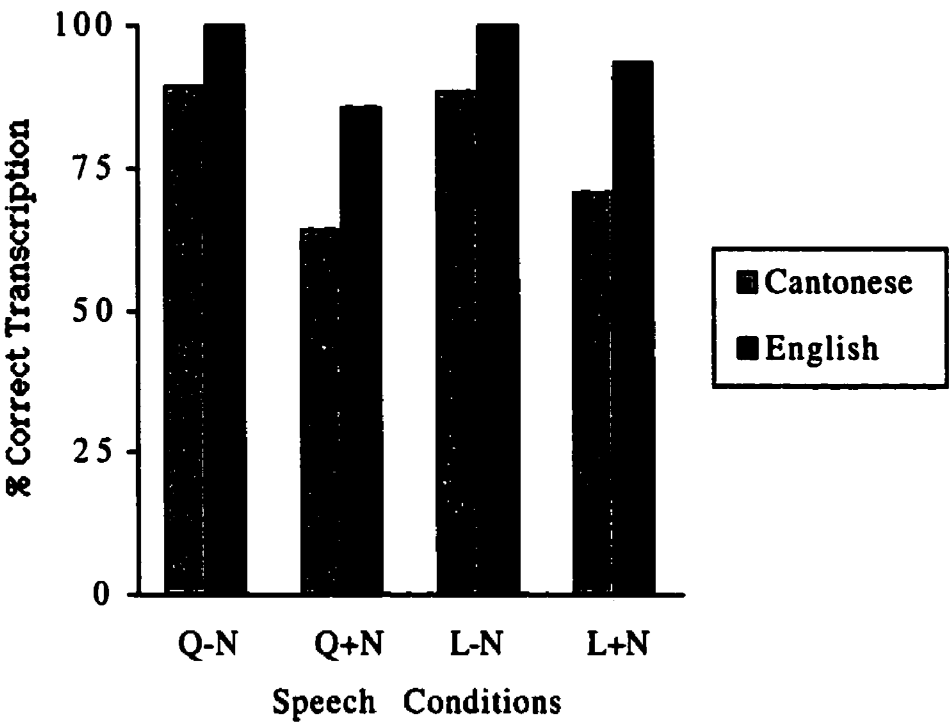


FIGURE 3-7

The effect of the four speech conditions on transcription (content word match) scores for the 12 native Cantonese speakers.

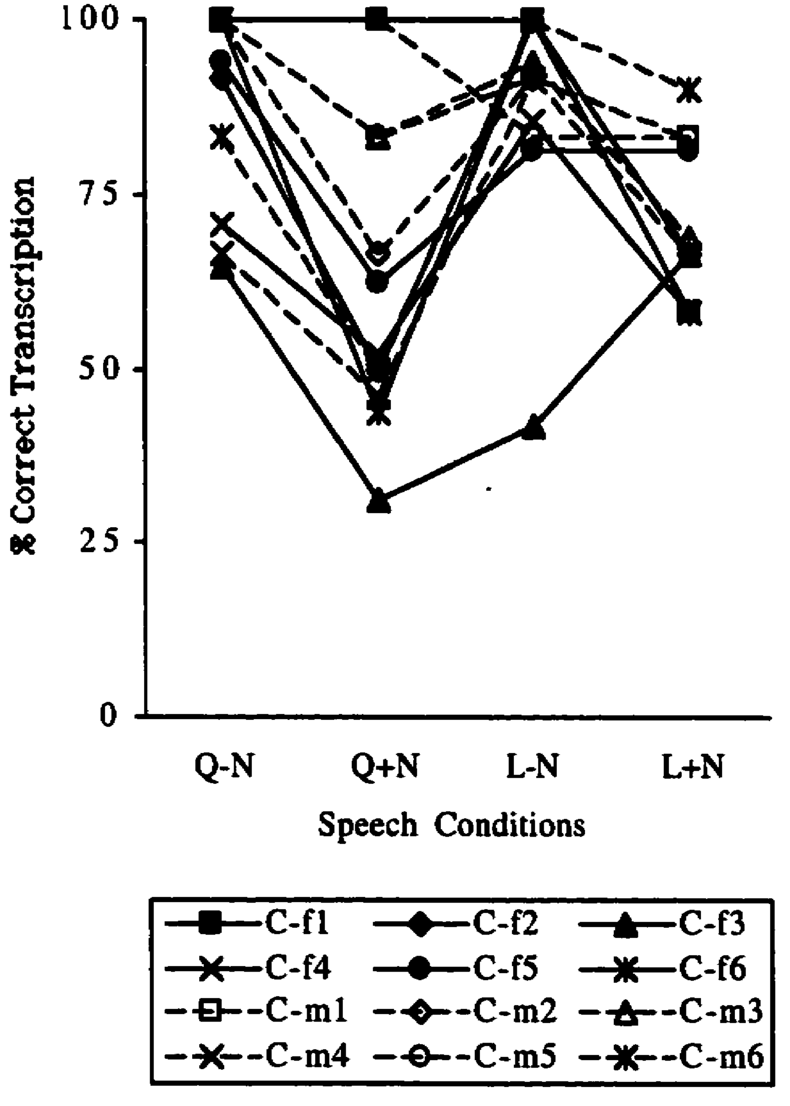


FIGURE 3-8

The effect of the four speech conditions on transcription (content word match) scores for the 12 native English speakers.

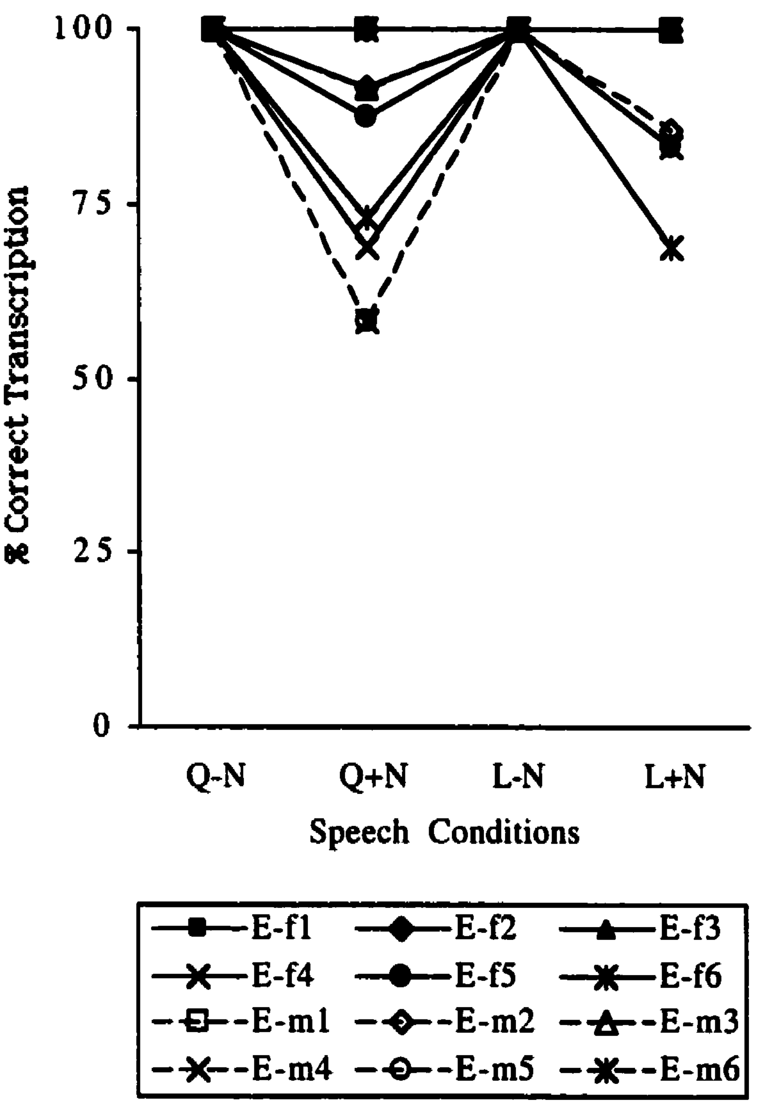


FIGURE 3-9

Percent correct verification scores for the native English and native Cantonese utterances presented under the four speech conditions.

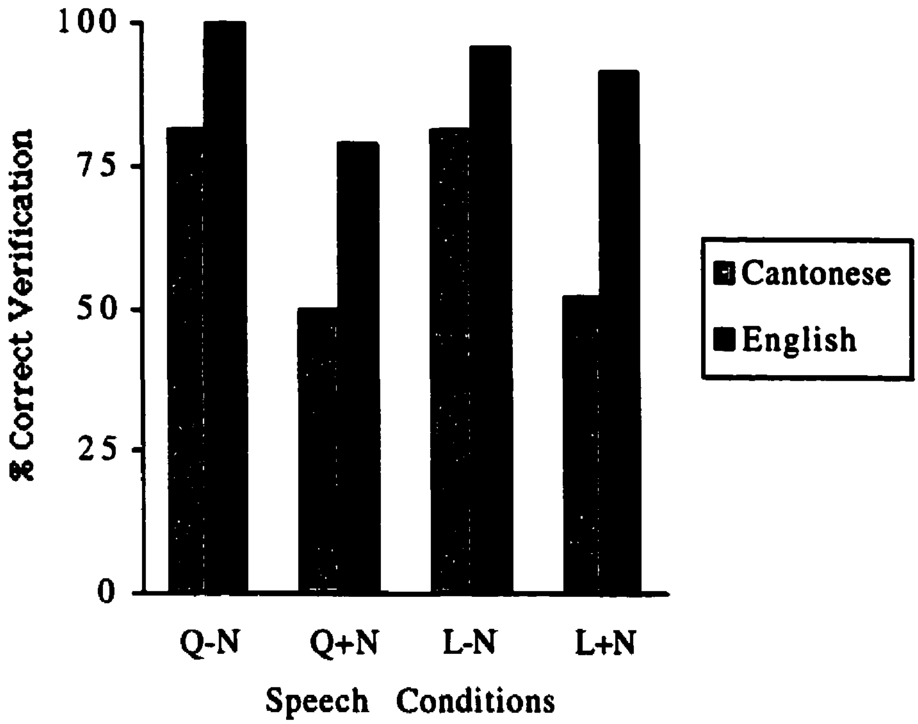


FIGURE 3-10

Mean foreign-accent ratings of the native English and native Cantonese speakers' utterances presented under the four speech conditions on a nine-point scale, where 1= no accent and 9= strong accent.

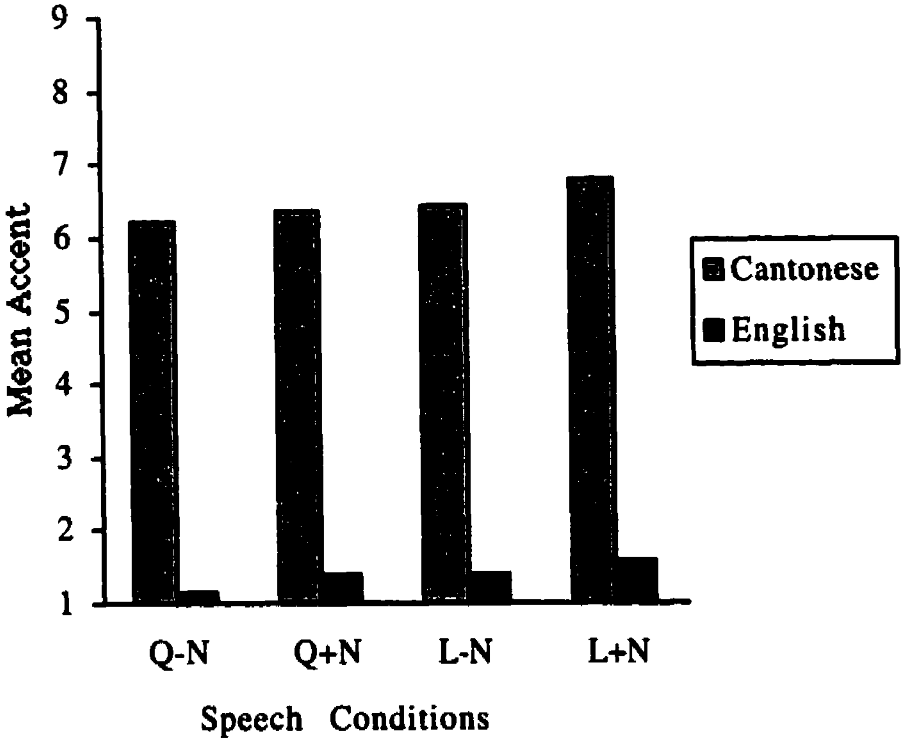
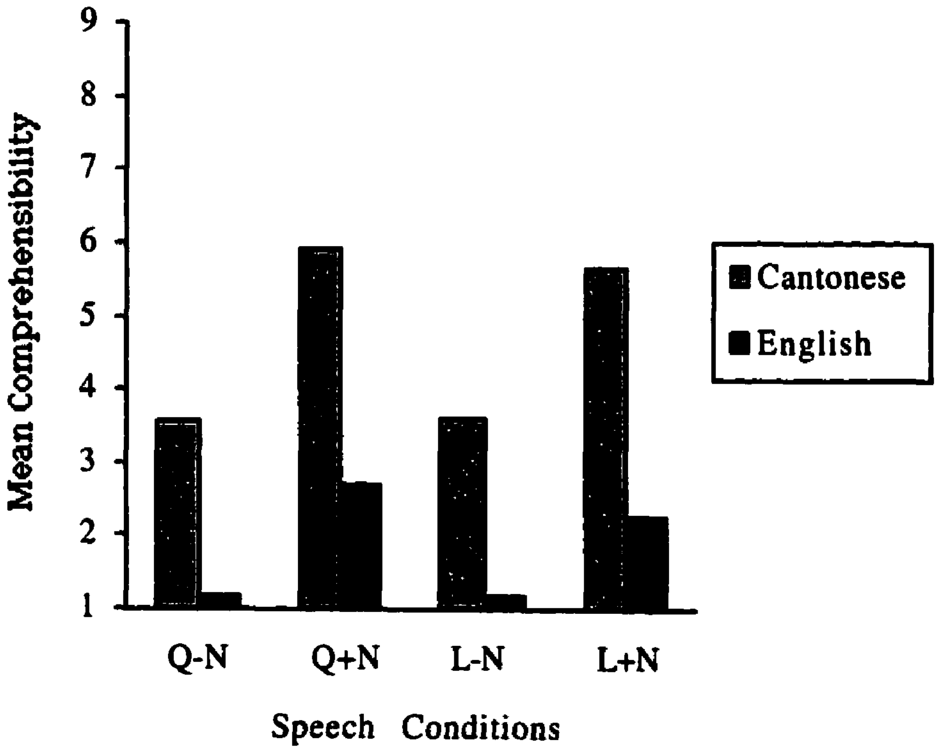


FIGURE 3-11

Mean comprehensibility ratings of the native English and native Cantonese speakers' utterances presented under the four speech conditions on a nine-point scale, where 1= not difficult to understand and 9= very difficult to understand.



CHAPTER IV

GENERAL DISCUSSION AND CONCLUSIONS

Despite the extensive literature on the Lombard reflex over the past decades, relatively little is known about the effects of noise on the production and perception of nonnative-produced English. The present study examined the role of the Lombard effect in the production and perception of Cantonese-accented English. This research represents a first attempt to report any change in speaking rate and sentential fundamental frequency values in simple, short English declarative sentences produced by native Hong Kong Cantonese speakers in the presence of noise. In addition, the perception experiment explores difference between normal and Lombard speech presented in noise-free and noisy conditions in terms of intelligibility, perceived accentedness and comprehensibility using standard procedures that are commonly employed in second language speech studies.

Role of the Lombard reflex in the production of Cantonese-accented English

In producing the English sentences, the Cantonese speakers exhibited one, but not both, of characteristics of Lombard speech examined in this experiment. In the presence of masking noise at 70 dB, the Cantonese speakers, as expected, exhibited higher mean sentential fundamental frequency and Fo range. Nonetheless, their speaking rates were more or less the same in both speaking conditions, whereas the native English reduced their speaking rates in noise. As already mentioned in Chapter 2, the Cantonese speakers, like other L2 learners of English, typically speak slower than do native English speakers in ordinary conditions. Their failure to slow down their rates in noisy conditions may be that the Cantonese speakers may not have perceived any advantage (e.g., in terms of producing more intelligible utterances) in reducing the speaking rates to a greater degree. The findings reported in this experiment bring out an interesting issue about speech production or communication under noisy conditions. Namely, the extent of the effects of the Lombard reflex on L1 production may not necessarily be the same on L2 production. However, only two characteristics have been studied in this experiment. It is worthwhile to look at other characteristics of the Lombard reflex, such as intensity levels, formant center frequencies (F1 and F2), durations and amplitudes of individual consonants and vowels, as mentioned in the literature (Junqua, 1996; Van Summers et al., 1988), before any conclusive statement can be made.

Role of the Lombard reflex in the perception of Cantonese-accented English

As in previous studies, the findings of the present research show that, in general, speech presented in noise is less well perceived than is noise-free speech. The masking noise was found to have deleterious effects on the intelligibility and perceived comprehensibility of normal and Lombard speech produced by the Cantonese speakers. This is similar to Munro's (1998) findings in that the addition of cafeteria noise to sentences was found to have a stronger effect on verification and transcription scores for his Mandarin speakers of English. In contrast, the perceived foreign-accentedness of Cantonese-accented English was found to be almost the same throughout all of the four speech conditions. There were no significant differences in listeners' ratings whether nonnative normal or Lombard speech was presented in noise or in quiet conditions. This finding implies that to native English-speaking listeners, the degree of perceived foreign-accent in English spoken by nonnative speakers appears not to be affected by whether the conversation takes place in a quiet environment (e.g., quiet classroom) or under nonideal listening conditions (e.g., in a noisy rendezvous or via the telephone). Moreover, the Lombard reflex is unlikely to have any impact on reducing the degree of accentedness in English sentences produced by the Cantonese speakers, despite its effectiveness in improving the intelligibility of speech in noisy conditions.

Not surprisingly, the noisy Lombard speech produced by the Cantonese speakers was found to be more intelligible than the noisy normal speech using the sentence-transcription task. The improvement in intelligibility is likely due in part to the higher sentential fundamental frequency values exhibited in Lombard speech. As mentioned earlier, an increase in Fo values is probably the result of an increase in voice levels (Bond et al., 1989). Of course, other aforementioned characteristics of Lombard speech may also be at work.

In contrast, this was not the case for the sentence-verification task. In the latter, Lombard speech was not correctly verified more frequently than normal speech in noisy conditions. The discrepancy in patterns for the intelligibility of Lombard and normal speech may be due to the different nature of the two tasks involved. In the orthographic transcription task, what the listeners need to do is simply to write down as many words as they can, upon listening to a stimulus sentence. They do not have to give much thought to verify its truth value. However, in the verification task, they have to process the meaning of the sentence to determine whether it is true or false. It may be that the sentence-verification task is a more holistic approach that is closer to "real-world" comprehension than is the sentence-transcription task. In verifying the truth value, the listeners do not have to catch every word (e.g., function word) in the sentence in making a decision. In addition, as stated by Pisoni and Dedina (1986), sentence verification is a less sensitive task than is orthographic transcription. The verification measure appears to be a coarse evaluation of intelligibility that may not be able to reveal small differences between nonnative Lombard and normal speech presented in noisy conditions at a constant S/N ratio.

Lombard speech has often been reported, in the present and previous research, to be more intelligible than normal speech under noisy conditions. Although a comparison of the difference in intelligibility between the two types of speech presented without masking noise has been neither carried out nor discussed, it is not unreasonable to expect that Lombard speech is also more intelligible than normal speech when presented in quiet conditions. Interestingly, for both groups of speakers in the current study, their conversation-like speech was found to be as intelligible and comprehensible as the Lombard speech when presented without masking noise. For each of the two groups of speakers, the scores or ratings in the condition of Q-N were more or less the same as those of the L-N condition. Since only two parameters have been examined in the production task here, no solid explanation for this phenomenon can be offered at this stage. It seems that the values of speaking rates and Fo displayed in Lombard speech, compared with those exhibited in normal utterances, do not contribute to any significant differences along the three dimensions, when both types of speech are perceived by native English listeners in noise-free conditions. There is room for further research in this issue.

Limitations and future directions

To the best of the author's knowledge, this research is the first of its kind not only to record nonnative speech samples in the presence of masking noise, but also to present both Lombard and normal speech in quiet and noisy conditions to native English listeners for evaluations of intelligibility, perceived foreign-accentedness and comprehensibility.

Nevertheless, some limitations of the present study should be identified. First of all, Van Summers et al. (1988) suggest that the effects of masking noise seem to be greater in speech tasks involving interaction or communication than in those with no external feedback (e.g., from an interlocutor). In this experiment, individual speaker-participants and the experimenter were seated across from one another at a distance of 2m. The speakers were requested to follow the instructions given by the experimenter, and to produce the sentences in a conversation-like manner similar to the way they talked in daily situations throughout the recordings. The purpose of using these strategies was to make the recording conditions more realistic when compared with those used in previous studies. However, spontaneous speech or even real communication, rather than a simple sentence-reading task used in this study, also need to be explored. As suggested by Junqua (1996), a consideration of these kinds of speech tasks may deepen our understanding of the Lombard reflex phenomenon.

In addition, only one group of L2 speakers (Cantonese speakers of English from Hong Kong) has been considered here. As a result, the findings of this research cannot be generalized too far. The listeners in this study may have benefited from familiarity with the particular accent used as the nonnative speakers all come from the same L1 background. So that the effects of the Lombard reflex on nonnative English speakers can be generalized, research with L2 speakers of other native languages needs to be carried out.

The present study focused on two characteristics of the Lombard reflex, speaking rate and fundamental frequency values, and identified them as interesting factors for exploring the effects of noise on nonnative-produced English. As already mentioned in Chapter I, an increase in intensity levels in speech is the major characteristic of the Lombard reflex. Because of lack of availability of equipment, however, this parameter has not been examined in this experiment. In addition, as in Van Summers et al. (1988), the overall RMS amplitudes of all the stimulus sentences for presentation to the listeners were equated. In fact, it would be valuable to examine any difference in voice levels in utterances spoken by the two groups of speakers in quiet and in noisy conditions in future.

Lane and Tranel (1971) state that the phenomenon of the Lombard reflex is important to research in speech communication in noise. As mentioned earlier, the phenomenon of the Lombard reflex may vary as a function of the type of noise used. In current research, only cafeteria-like masking noise was employed. Future studies of the Lombard reflex using different kinds of noise (e.g., car noise, or computer room noise) may contribute a better understanding of the effects of real-world noise on the production and perception of native- or nonnative-produced speech.

Appendix 1

Language Background Questionnaire -- Cantonese Speakers/Listeners

Today's Date: _____________ Participant Code:__________

1. Do you have normal hearing? Y N

2. What is your date of birth?_____/_____/_____(DD/MM/YY)

3. Were you born and raised in Hong Kong? Y N

4. Is your first language Cantonese? Y N

5. Did you complete at least part of Form Five in Hong Kong? Y N

6. What was the highest level of education that you attained in Hong Kong?___________________________

7. At what age did you begin studying English in school?_________

8. Were any of your teachers in Hong Kong native speakers of English? Y N

   If answer to (8) is Yes: (a) How many? ____________

   (b) For how long?____________

9. When did you arrive in Canada?

   _____/_____/_____(DD/MM/YY)

10. For how many years (total) did you study English before coming to Canada? ______

11. List all other places where you have lived for more than 6 months:

Where	When	How long

12. Have you ever been away from Canada for more than 1 month since you arrived? Where and for how long?

13. Have you ever taken any ESL classes in Canada? If so, for how long?

14. What other languages have you studied in school or university?

Language	For how long	How proficient are you?
__________	____________	______________
__________	____________	______________
__________	____________	______________

15. Do you speak any other languages or Chinese dialects? Which ones?

16. Have you ever taken a special course to improve your pronunciation? If so, how long (in weeks) was it?

17. On an average day, what percentage of the time do you use English for communication?

__________%

18. How often do you speak English when you are at home?

__________%

19. How often do you speak English when you are not at home?

__________%

20. How often do you speak English when you are socializing with your friends?

__________%

Appendix 2
Language Background Questionnaire-English Speakers/Listeners

Today's Date: _____________ Participant Code: _________

1. Do you have normal hearing? Y N

2. What is your date of birth?____/____/____(DD/MM/YY)

3. Were you born and raised in western Canada? Y N

If so, where? __________

4. Is your first language English? Y N

5. What other languages have you studied in school or university?

Language	For how long	How proficient are you?
________	___________	________________
________	___________	________________
________	___________	________________
________	___________	________________

6. Do you speak any other languages? Which ones?

Appendix 3

List of True (1-24) and False (25-28) sentences with the number of words.

Number	Sentence	Number of Words
1	You can *take pictures* with a *camera*	7
2	*Bread* is *made* with *flour*	5
3	An *engine* is a *part* of a *ship*	8
4	A *tiger* is *bigger* than a *cat*	7
5	You can *tell time* with a *watch*	7
6	*India* is in *Asia*	4
7	*People* can *ride camels* in the *desert*	7
8	A *pigeon* is a *kind* of *bird*	7
9	You can *borrow* a *book* from a *library*	8
10	*Water* and *sunlight* are *essential* to *flowers*	7
11	*Seven* is an *odd number*	5
12	*Christmas* is in *December*	4
13	*Elephants* are *big animals*	4
14	*Hot* and *cold* are *opposites*	5
15	*Exercise* is *good* for your *health*	6
16	*Japan* is a *wealthy country*	5
17	*Ships travel* on the *water*	5
18	Some *people keep dogs* as *pets*	6
19	*Young children* can be very *noisy*	6
20	Some *roses* have a *beautiful smell*	6
21	*Hungry cats like* to *chase mice*	6
22	*Italy* is a *country* in *Europe*	6
23	*Red* and *green* are *colors*	5
24	*Gold* is a *valuable metal*	5

Number	Sentence	Number of Words
25	*California* is in *Russia*	4
26	*Christmas* is in *September*	4
27	Most *cats like* to *read magazines*	6
28	In *summer* the *sun* is *blue*	6
29	A *spider* is *bigger* than a *cat*	7
30	Some *chickens live* on the *moon*	6
31	*Butterflies need batteries* to *fly*	5
32	You can *buy vegetables* at the *bank*	7
33	It's *good* to *have stones* for *breakfast*	7
34	All *scientists have three brains*	5
35	*People brush* their *teeth* with a *telephone*	7
36	You can *tell time* with a *kettle*	7
37	*Gasoline* is an *excellent drink*	5
38	The *sun always sets* in the *north*	7
39	The *inside* of an *egg* is *blue*	7
40	*August* is a *winter month*	5
41	It *always snows* in *July*	5
42	Most *people wear hats* on their *feet*	7
43	The *stars come* out in the *day*	7
44	*Wednesday* is the *first day* of the *week*	8
45	All *men* can *have babies*	5
46	All *dogs have fifteen legs*	5
47	*People eat* through their *noses*	5
48	A *monkey* is a *kind* of *bird*	7

Appendix 4

Record Sheet (T/F: List A)

Participant's Code:__________ Today's Date:__________

Condition	Sequence
Quiet	1st / 2nd
Noise	1st / 2nd

On the No., mark √(for acceptable utterance), or O (for unacceptable utterance to be repeated at the end of each list)

List A 1

13. Elephants are big animals.
23. Red and green are colors.
15. Exercise is good for your health.
43. The stars come out in the day.
16. Japan is a wealthy country.
13. Elephants are big animals.
18. Some people keep dogs as pets.
1. You can take pictures with a camera.
28. In summer the sun is blue.
16. Japan is a wealthy country.

List A 2

14. Hot and cold are opposites.
39. The inside of an egg is blue.
2. Bread is made with flour.
8. A pigeon is a kind of bird.
5. You can tell time with a watch.
14. Hot and cold are opposites.
29. A spider is bigger than a cat.
41. It always snows in July.
30. Some chickens live on the moon.
5. You can tell time with a watch.

List A 3

19. Young children can be very noisy.
47. People eat through their noses.
42. Most people wear hats on their feet.
4. A tiger is bigger than a cat.
31. Butterflies need batteries to fly.
19. Young children can be very noisy.
12. Christmas is in December.
48. A monkey is a kind of bird.
11. Seven is an odd number.
31. Butterflies need batteries to fly.

List A 4

33. It's good to have stones for breakfast.
21. Hungry cats like to chase mice.
9. You can borrow a book from a library.
22. Italy is a country in Europe.
7. People can ride camels in the desert.
33. It's good to have stones for breakfast.
17. Ships travel on the water.
40. August is a winter month.
3. An engine is a part of a ship.
7. People can ride camels in the desert.

List A 5

44. Wednesday is the first day of the week.
45. All men can have babies.
10. Water and sunlight are essential to flowers.
27. Most cats like to read magazines.
24. Gold is a valuable metal.
44. Wednesday is the first day of the week.
35. People brush their teeth with a telephone.
6. India is in Asia.
36. You can tell time with a kettle.
24. Gold is a valuable metal.

List A 6

46. All dogs have fifteen legs.
38. The sun always sets in the north.
34. All scientists have three brains.
26. Christmas is in September.
32. You can buy vegetables at the bank.
46. All dogs have fifteen legs.
25. California is in Russia.
37. Gasoline is an excellent drink.
20. Some roses have a beautiful smell.
32. You can buy vegetables at the b

Appendix 5
Stimulus Lists

	Group One								Group One							
	Q- N		Q+ N		L- N		L+ N		Q- N		Q+ N		L- N		L+ N	
Speaker	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F
C-m1	12	34	12	34	12	34	12	34	21	29	21	29	21	29	21	29
C-m2	23	37	23	37	23	37	23	37	19	36	19	36	19	36	19	36
C-m3	17	33	17	33	17	33	17	33	4	40	4	40	4	40	4	40
C-m4	9	26	9	26	9	26	9	26	17	37	17	37	17	37	17	37
C-m5	15	36	15	36	15	36	15	36	23	32	23	32	23	32	23	32
C-m6	21	28	21	28	21	28	21	28	13	42	13	42	13	42	13	42
C-f1	3	29	3	29	3	29	3	29	22	25	22	25	22	25	22	25
C-f2	8	39	8	39	8	39	8	39	2	28	2	28	2	28	2	28
C-f3	14	38	14	38	14	38	14	38	8	33	8	33	8	33	8	33
C-f4	18	27	18	27	18	27	18	27	14	30	14	30	14	30	14	30
C-f5	1	45	1	45	1	45	1	45	18	35	18	35	18	35	18	35
C-f6	22	35	22	35	22	35	22	35	6	45	6	45	6	45	6	45
E-m1	7	25	7	25	7	25	7	25	16	46	16	46	16	46	16	46
E-m2	11	32	11	32	11	32	11	32	5	27	5	27	5	27	5	27
E-m3	13	41	13	41	13	41	13	41	15	47	15	47	15	47	15	47
E-m4	20	31	20	31	20	31	20	31	24	39	24	39	24	39	24	39
E-m5	2	47	2	47	2	47	2	47	12	34	12	34	12	34	12	34
E-m6	24	42	24	42	24	42	24	42	11	43	11	43	11	43	11	43
E-f1	19	40	19	40	19	40	19	40	1	26	1	26	1	26	1	26
E-f2	6	48	6	48	6	48	6	48	9	44	9	44	9	44	9	44
E-f3	10	30	10	30	10	30	10	30	20	48	20	48	20	48	20	48
E-f4	4	44	4	44	4	44	4	44	10	41	10	41	10	41	10	41
E-f5	5	46	5	46	5	46	5	46	3	31	3	31	3	31	3	31
E-f6	16	43	16	43	16	43	16	43	7	38	7	38	7	38	7	38

Shaded portions are sentences masked with noise.

Appendix 6

RMS amplitude values (dB) and attenuating factors of sentences in Q-N of Group One.

Sentence Number	RMS (dB)	raw RMS value	Attenuating Factor (AF)	AF in Percent (%)
1	51.91	393.79	80.30	80
2	53.18	456.27	69.31	69
3	53.43	469.23	67.39	67
4	46.02	199.96	158.14	158
5	49.38	294.39	107.42	107
6	49.97	314.99	100.39	100
7	52.11	403.10	78.45	78
8	47.54	238.31	132.69	133
9	52.40	416.78	75.87	76
10	49.61	302.19	104.64	105
11	50.87	349.61	90.45	90
12	54.24	515.26	61.37	61
13	52.77	435.19	72.67	73
14	52.05	400.35	78.99	79
15	45.03	178.48	177.17	177
16	48.81	275.75	114.68	115
17	46.60	213.89	147.85	148
18	45.45	187.20	168.93	169
19	51.88	392.85	80.50	80
20	48.96	280.70	112.66	113
21	45.02	178.29	177.37	177
22	53.69	483.51	65.40	65
23	51.29	366.80	86.21	86
24	54.61	537.45	58.84	59
25	51.40	371.57	85.11	85
26	46.72	216.73	145.91	146
27	47.93	249.18	126.91	127
28	47.54	238.36	132.67	133
29	51.37	370.10	85.44	85
30	49.64	303.36	104.24	104
31	52.27	410.58	77.02	77
32	51.44	373.13	84.75	85
33	44.41	166.21	190.26	190
34	54.64	539.53	58.61	59
35	45.21	182.23	173.54	174
36	52.45	419.11	75.45	75
37	51.26	365.78	86.45	86
38	47.97	250.31	126.34	126
39	53.02	447.90	70.60	71
40	50.93	352.03	89.83	90
41	52.18	406.41	77.81	78
42	49.51	298.85	105.82	106
43	51.86	391.62	80.75	81
44	44.71	172.02	183.83	184
45	52.60	426.56	74.13	74
46	45.11	180.01	175.68	176
47	51.82	389.78	81.13	81
48	51.02	355.68	88.91	89

Mean (dB) 50.08

Appendix 7

RMS amplitude values (dB) and attenuating factors for sentences in L-N of Group One.

Sentence Number	RMS (dB)	raw RMS value	Attenuating Factor (AF)	AF in Percent (%)
1	51.57	378.93	83.45	83
2	49.53	299.42	105.61	106
3	55.79	616.05	51.33	51
4	50.75	344.55	91.78	92
5	46.57	212.96	148.49	148
6	54.24	515.52	61.34	61
7	48.45	264.63	119.50	119
8	50.22	324.39	97.48	97
9	55.34	584.70	54.08	54
10	48.75	273.90	115.45	115
11	53.89	494.61	63.93	64
12	51.28	366.46	86.29	86
13	52.51	422.25	74.89	75
14	53.57	477.22	66.27	66
15	45.22	182.33	173.44	173
16	49.56	300.55	105.22	105
17	47.54	238.11	132.81	133
18	44.25	163.07	193.93	194
19	49.28	291.05	108.65	109
20	48.72	272.97	115.85	116
21	44.20	162.27	194.88	195
22	52.75	433.82	72.89	73
23	51.32	367.98	85.94	86
24	52.35	414.28	76.33	76
25	54.00	501.46	63.06	63
26	46.86	220.37	143.50	143
27	47.24	230.16	137.40	137
28	48.80	275.52	114.78	115
29	49.93	313.80	100.77	101
30	48.72	272.88	115.89	116
31	51.09	358.62	88.18	88
32	52.09	402.11	78.64	79
33	44.85	174.88	180.83	181
34	53.83	491.42	64.35	64
35	47.02	224.47	140.88	141
36	49.60	301.92	104.74	105
37	48.49	265.70	119.02	119
38	52.43	418.27	75.60	76
39	51.02	355.83	88.87	89
40	49.00	281.90	112.18	112
41	50.76	345.21	91.60	92
42	48.54	267.40	118.26	118
43	50.96	353.28	89.51	90
44	48.81	275.71	114.70	115
45	51.35	369.19	85.65	86
46	49.51	298.88	105.81	106
47	49.21	288.60	109.57	110
48	53.22	458.35	68.99	69

Mean (dB) 50.10

Appendix 8

RMS amplitude values (dB) and attenuating factors of sentences in Q-N of Group Two.

Sentence Number	RMS (dB)	raw RMS value	Attenuating Factor (AF)	AF in Percent (%)
1	49.23	289.244	109.33	109
2	48.60	269.265	117.44	117
3	50.38	330.370	95.72	96
4	48.90	278.731	113.45	113
5	52.00	398.231	79.41	79
6	50.73	343.855	91.97	92
7	51.63	381.417	82.91	83
8	51.58	379.271	83.38	83
9	55.03	564.326	56.04	56
10	48.81	275.807	114.66	115
11	50.82	347.616	90.97	91
12	48.08	253.627	124.68	125
13	47.53	238.013	132.86	133
14	49.77	307.914	102.70	103
15	50.93	352.128	89.80	90
16	52.18	406.284	77.83	78
17	51.38	370.843	85.27	85
18	47.24	230.028	137.47	137
19	54.26	516.577	61.22	61
20	49.63	303.047	104.35	104
21	54.94	558.348	56.64	57
22	58.50	840.998	37.60	38
23	47.86	247.044	128.00	128
24	52.27	410.667	77.00	77
25	59.47	940.871	33.61	34
26	51.29	366.796	86.21	86
27	50.59	338.478	93.43	93
28	51.46	373.895	84.58	85
29	54.99	561.811	56.29	56
30	48.01	251.478	125.75	126
31	51.72	385.363	82.06	82
32	47.63	240.688	131.38	131
33	49.14	286.447	110.40	110
34	51.78	388.253	81.45	81
35	48.55	267.747	118.11	118
36	53.79	489.221	64.64	65
37	50.67	341.645	92.56	93
38	49.84	310.420	101.87	102
39	49.06	283.720	111.46	111
40	48.80	275.325	114.86	115
41	49.54	299.751	105.50	105
42	48.06	252.968	125.01	125
43	51.80	389.260	81.24	81
44	50.87	349.615	90.45	90
45	53.88	494.026	64.01	64
46	53.14	453.759	69.69	70
47	51.55	377.985	83.66	84
48	49.14	286.260	110.47	110

Mean (dB) 50.98

Appendix 9

RMS amplitude values (dB) and attenuating factors of sentences in L-N of Group Two.

Sentence Number	RMS (dB)	raw RMS value	Attenuating Factor (AF)	AF in Percent (%)
1	51.91	393.863	80.29	80
2	52.98	445.513	70.98	71
3	46.20	204.146	154.90	155
4	45.85	196.198	161.18	161
5	50.95	352.874	89.61	90
6	52.70	431.425	73.30	73
7	48.80	275.572	114.75	115
8	52.00	397.965	79.46	79
9	50.80	346.685	91.21	91
10	48.24	258.312	122.42	122
11	51.23	364.280	86.81	87
12	48.30	259.902	121.67	122
13	51.28	366.357	86.32	86
14	50.30	327.439	96.58	97
15	47.34	232.932	135.76	136
16	52.24	409.397	77.24	77
17	48.75	273.883	115.46	115
18	50.24	325.042	97.29	97
19	49.42	295.920	106.86	107
20	46.27	205.769	153.68	154
21	52.79	435.833	72.56	73
22	52.70	431.291	73.32	73
23	51.89	393.034	80.46	80
24	51.87	392.008	80.67	81
25	53.08	450.749	70.16	70
26	52.01	398.382	79.38	79
27	50.35	329.416	96.00	96
28	49.94	313.942	100.73	101
29	52.43	418.182	75.62	76
30	48.11	254.302	124.35	124
31	47.28	231.079	136.85	137
32	49.13	285.970	110.58	111
33	48.15	255.494	123.77	124
34	46.40	208.836	151.42	151
35	48.31	260.181	121.54	122
36	48.56	267.988	118.00	118
37	47.10	226.548	139.59	140
38	50.54	336.624	93.94	94
39	53.34	464.478	68.08	68
40	46.85	220.146	143.64	144
41	49.13	286.121	110.52	111
42	45.19	181.823	173.92	174
43	51.45	373.788	84.60	85
44	50.41	331.326	95.44	95
45	51.25	365.094	86.62	87
46	53.46	470.874	67.16	67
47	49.71	305.876	103.38	103
48	49.57	301.030	105.05	105

Mean (dB) 49.93

Appendix 10

Equated RMS amplitude values (dB) of sentences in Q-N of Group One.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	71.80	15.56	49.95
2	71.39	19.99	49.94
3	70.81	16.90	49.88
4	76.36	19.99	50.02
5	71.70	15.56	49.97
6	65.99	19.08	49.97
7	67.74	21.58	49.93
8	74.45	21.58	50.03
9	74.74	12.04	49.99
10	74.91	13.97	50.03
11	74.36	22.27	49.95
12	72.31	20.82	49.92
13	70.00	22.92	50.02
14	70.78	13.97	49.99
15	75.40	21.58	50.02
16	75.60	19.08	50.02
17	74.66	22.27	50.03
18	78.22	27.60	50.03
19	67.04	19.08	49.93
20	74.98	20.82	50.03
21	76.90	22.27	50.02
22	65.48	20.82	49.92
23	73.04	13.97	49.96
24	69.21	16.90	49.98
25	70.19	19.08	49.97
26	77.11	19.08	50.03
27	76.15	24.08	50.01
28	77.01	16.90	50.04
29	74.35	16.90	49.94
30	74.93	18.06	49.98
31	72.05	24.08	49.98
32	69.18	22.27	50.01
33	78.00	24.60	50.04
34	73.46	21.58	50.03
35	79.89	22.92	50.06
36	67.98	24.08	49.93
37	74.83	13.97	49.95
38	76.45	19.99	49.99
39	69.01	22.92	50.03
40	76.87	21.58	50.00
41	68.80	18.06	50.00
42	72.69	22.27	50.02
43	67.85	22.27	50.02
44	75.62	19.08	50.06
45	67.74	24.60	49.97
46	77.35	20.82	50.06
47	69.21	24.60	49.97
48	68.35	22.27	50.00
Maximum (dB)			50.06
Minimum (dB)			49.88
Percentage Difference (%)			0.36
Mean (dB)			49.99
Standard Deviation			0.04

Appendix 11

Equated RMS amplitude values (dB) of sentences in L-N of Group One.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	71.06	13.97	49.94
2	70.91	21.58	50.03
3	72.38	15.56	49.89
4	74.69	18.06	50.01
5	73.82	24.60	49.99
6	68.73	16.90	49.93
7	71.07	12.04	50.08
8	71.88	18.06	49.93
9	71.60	12.04	49.94
10	75.94	21.58	49.97
11	69.51	21.58	49.98
12	73.01	24.60	49.96
13	69.22	16.90	50.00
14	71.76	15.56	49.92
15	77.27	23.52	50.02
16	71.51	20.82	49.98
17	74.99	21.58	49.95
18	80.10	22.92	50.04
19	71.23	13.97	50.03
20	75.12	20.82	50.02
21	80.82	24.60	50.05
22	69.61	19.08	50.00
23	72.31	15.56	50.00
24	70.50	16.90	49.93
25	68.83	21.58	49.96
26	75.72	22.27	49.98
27	77.84	22.92	49.99
28	76.87	19.99	50.03
29	75.13	18.06	50.04
30	75.08	23.52	50.01
31	71.27	18.06	49.97
32	69.05	22.92	50.03
33	76.82	24.60	50.05
34	74.62	22.27	49.92
35	80.45	21.58	50.03
36	74.46	16.90	50.02
37	79.71	21.58	50.01
38	72.01	16.90	50.02
39	72.15	19.99	50.00
40	77.05	18.06	50.00
41	73.00	16.90	50.03
42	75.16	19.99	49.98
43	72.37	19.99	50.04
44	72.70	21.58	50.02
45	71.61	13.97	50.03
46	75.03	21.58	50.02
47	69.47	19.08	50.04
48	70.62	13.97	49.96

Maximum (dB)	50.08
Minimum (dB)	49.89
Percentage Difference	0.37
Mean (dB)	50.00
Standard Deviation	0.04

Appendix 12

Equated RMS amplitude values (dB) of sentences in Q-N of Group Two.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	73.59	18.06	49.94
2	76.82	21.58	49.98
3	70.48	19.08	50.02
4	74.12	19.08	49.97
5	68.84	24.08	49.95
6	72.60	18.06	50.00
7	72.37	18.06	49.99
8	76.36	15.56	49.95
9	70.38	12.04	49.95
10	76.22	15.56	50.02
11	71.77	24.60	49.99
12	76.47	22.27	50.03
13	79.35	21.58	50.02
14	78.07	19.99	50.03
15	70.43	13.97	50.01
16	73.99	24.08	49.99
17	74.81	16.90	49.96
18	73.33	24.08	49.98
19	70.10	15.56	49.93
20	75.86	18.06	49.97
21	72.61	19.99	50.03
22	67.89	18.06	50.05
23	77.98	19.99	50.01
24	68.81	19.99	49.98
25	68.38	16.90	50.06
26	74.16	15.56	49.97
27	71.90	25.57	49.95
28	71.08	15.56	50.03
29	74.40	19.99	49.91
30	76.00	24.60	50.03
31	71.76	15.56	49.98
32	75.48	21.58	49.99
33	78.37	19.99	49.97
34	74.19	15.56	49.93
35	73.88	18.06	49.99
36	71.68	16.90	50.02
37	72.72	13.97	50.03
38	76.44	23.52	50.01
39	73.10	19.08	49.96
40	76.46	22.27	50.01
41	75.79	16.90	49.95
42	77.60	18.06	50.01
43	69.72	22.92	49.96
44	72.71	20.82	49.95
45	69.36	13.97	49.97
46	70.90	22.27	50.00
47	70.48	19.99	50.03
48	74.63	16.90	49.96

Maximum (dB)	50.06
Minimum (dB)	49.91
Percentage Difference	0.28
Mean (dB)	49.99
Standard Deviation	0.03

Appendix 13

Equated RMS amplitude values (dB) of sentences in L-N of Group Two.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	70.80	22.92	49.95
2	68.88	20.82	49.98
3	75.40	26.02	50.03
4	74.52	24.08	50.02
5	73.23	22.92	50.04
6	65.53	23.52	49.95
7	73.23	22.27	49.99
8	67.52	20.82	49.93
9	67.71	16.90	49.96
10	72.01	19.99	49.98
11	71.35	25.10	50.01
12	72.65	27.23	50.03
13	71.23	16.90	49.96
14	74.10	18.06	50.02
15	73.25	21.58	50.03
16	71.76	24.08	49.95
17	73.34	19.08	49.97
18	73.59	18.06	49.96
19	70.21	15.56	49.97
20	76.78	21.58	50.04
21	74.77	22.27	50.03
22	68.38	20.82	49.94
23	70.04	16.90	49.93
24	68.11	19.99	50.01
25	68.11	18.06	49.96
26	71.45	25.10	49.95
27	73.58	24.60	49.99
28	73.74	18.06	50.02
29	72.26	19.08	50.02
30	69.80	25.10	49.98
31	74.73	19.08	50.03
32	73.72	16.90	50.04
33	72.33	19.08	50.04
34	76.31	22.92	50.00
35	74.95	20.82	50.04
36	74.60	19.99	50.00
37	72.09	24.08	50.04
38	73.45	18.06	50.00
39	70.14	22.92	49.97
40	76.15	23.52	50.04
41	73.65	16.90	50.04
42	79.05	19.99	50.07
43	69.72	22.27	50.03
44	73.73	19.08	49.95
45	70.27	19.99	50.03
46	73.17	21.58	49.95
47	72.06	22.27	49.97
48	73.15	16.90	50.00

Maximum (dB)	50.07
Minimum (dB)	49.93
Percentage Difference (%)	0.29
Mean (dB)	50.00
Standard Deviation	0.04

Appendix 14

Equated RMS amplitude values (dB) of sentences in Q+N of Group One.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	65.03	38.58	49.88
2	64.90	39.73	49.92
3	64.64	39.36	49.93
4	69.49	39.27	49.90
5	64.70	39.36	49.96
6	60.19	40.66	49.93
7	62.78	41.28	50.03
8	67.08	38.58	50.05
9	67.33	38.68	49.94
10	68.03	39.55	49.94
11	68.60	40.66	50.01
12	66.44	40.58	50.05
13	63.97	40.50	49.91
14	64.31	39.64	49.92
15	68.48	39.27	49.92
16	69.51	40.42	49.91
17	67.89	39.73	50.06
18	71.63	40.08	49.92
19	61.73	41.28	50.04
20	69.03	40.08	49.98
21	70.17	39.36	49.90
22	60.01	40.50	50.05
23	66.33	39.73	49.97
24	62.86	39.82	49.98
25	65.50	41.58	49.93
26	70.62	39.18	50.05
27	69.77	39.82	49.99
28	70.32	39.73	49.95
29	67.05	39.18	50.05
30	69.31	40.82	49.95
31	65.93	40.08	50.06
32	63.38	40.66	50.01
33	70.46	38.38	50.06
34	67.16	39.55	49.93
35	72.56	38.98	50.07
36	62.71	41.06	49.98
37	67.87	38.88	50.00
38	69.40	39.46	49.96
39	62.62	39.64	50.05
40	70.74	40.25	50.01
41	63.11	40.82	49.96
42	67.24	40.98	50.05
43	62.03	40.58	49.97
44	68.94	39.73	50.01
45	61.67	40.42	50.01
46	70.26	39.36	49.92
47	62.91	40.17	49.92
48	62.02	40.17	49.99

Maximum (dB)	50.07
Minimum (dB)	49.88
Percenage Difference (%)	0.38
Mean (dB)	49.98
Standard Deviation	0.06

Appendix 15

Equated RMS amplitude values (dB) of sentences in L+N of Group One.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	63.96	39.36	50.05
2	64.52	40.08	49.90
3	64.92	38.16	50.05
4	67.33	39.18	49.95
5	66.86	39.64	50.05
6	62.04	39.73	49.98
7	64.23	39.55	49.96
8	64.90	38.88	49.93
9	64.34	39.36	49.92
10	68.86	39.27	49.90
11	63.22	40.17	49.96
12	67.43	40.74	50.05
13	62.84	39.73	49.97
14	64.59	38.98	50.03
15	69.50	38.06	50.01
16	66.03	40.98	50.06
17	68.06	39.73	49.91
18	73.15	39.46	49.92
19	64.96	40.08	50.00
20	69.26	40.66	50.03
21	73.30	38.88	49.91
22	63.75	40.08	50.02
23	65.63	39.73	50.02
24	63.82	39.64	49.91
25	63.55	41.13	49.93
26	70.08	40.42	49.95
27	71.31	39.91	49.98
28	69.99	39.27	50.06
29	67.39	38.16	49.92
30	69.01	40.58	49.95
31	64.86	40.08	49.90
32	63.41	40.25	50.05
33	69.39	38.88	50.01
34	67.93	39.18	49.92
35	73.11	39.18	50.03
36	67.59	39.55	49.93
37	72.53	39.46	49.98
38	65.24	39.55	50.01
39	65.42	39.18	49.91
40	70.35	39.46	50.03
41	65.97	39.36	50.07
42	69.16	40.58	49.94
43	65.63	39.73	49.94
44	65.72	39.64	50.00
45	64.91	39.73	49.93
46	67.84	39.27	49.92
47	63.83	40.82	50.02
48	62.87	38.58	49.90

Maximum (dB)	50.07
Minimum (dB)	49.90
Percentage Difference (%)	0.35
Mean (dB)	49.97
Standard Deviation	0.05

Appendix 16

Equated RMS amplitude values (dB) of sentences in Q+N of Group Two.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	67.06	39.91	49.99
2	68.98	37.95	49.96
3	63.99	39.82	50.06
4	66.80	38.79	49.96
5	63.56	41.21	50.02
6	66.46	40.25	49.99
7	66.11	40.25	49.89
8	69.48	39.64	49.96
9	63.28	39.08	50.03
10	69.70	39.55	50.04
11	66.11	40.82	49.94
12	71.69	41.72	49.94
13	71.98	39.27	49.95
14	70.49	38.88	49.97
15	63.59	39.73	50.05
16	68.35	40.90	49.97
17	66.92	38.48	49.92
18	66.28	38.79	50.00
19	63.23	39.55	49.97
20	69.81	39.99	49.95
21	66.32	39.46	49.93
22	62.19	40.08	50.01
23	70.64	38.88	49.90
24	63.23	40.50	49.94
25	63.58	41.65	50.02
26	68.00	40.25	50.07
27	66.85	41.13	50.01
28	64.16	39.36	49.89
29	67.02	38.58	50.06
30	69.61	40.17	49.97
31	63.80	38.16	49.90
32	69.03	39.46	49.95
33	71.05	38.58	49.92
34	67.61	39.91	50.00
35	67.40	39.55	49.89
36	64.71	39.36	49.99
37	66.26	39.64	49.91
38	69.66	39.55	49.98
39	67.21	40.66	49.95
40	69.21	39.27	49.91
41	68.71	39.46	49.91
42	70.49	39.46	50.02
43	64.32	40.82	50.01
44	65.11	38.58	50.05
45	62.56	39.08	49.93
46	65.31	40.90	49.93
47	63.71	39.73	49.91
48	68.67	40.17	50.03

Maximum (dB)	50.07
Minimum (dB)	49.89
% Difference	0.35
Mean (dB)	49.97
Standard Deviation	0.05

Appendix 17

Equated RMS amplitude values (dB) of sentences in L+N of Group Two.

Sentence Number	Maximum RMS (dB)	Minimum RMS (dB)	Mean RMS (dB)
1	64.75	40.50	49.92
2	62.79	40.34	49.98
3	69.04	39.99	50.04
4	67.78	39.73	49.99
5	67.23	40.17	49.96
6	60.74	41.43	50.02
7	67.75	41.13	50.02
8	61.54	40.34	50.06
9	61.37	40.17	49.92
10	65.72	40.17	50.03
11	65.63	40.82	50.03
12	67.69	41.65	49.95
13	64.59	39.73	50.05
14	66.60	38.88	49.96
15	65.95	39.18	50.04
16	65.97	40.66	49.94
17	66.62	39.36	49.99
18	66.02	38.79	49.92
19	63.54	39.64	50.03
20	70.05	39.36	49.89
21	68.42	40.08	50.06
22	62.78	39.91	50.03
23	63.08	39.46	49.97
24	62.23	40.50	49.99
25	63.37	41.79	50.04
26	66.13	41.21	50.02
27	67.37	40.25	49.91
28	66.82	38.98	49.97
29	64.72	38.68	49.89
30	64.07	40.98	50.00
31	67.02	38.38	49.94
32	66.23	38.06	49.99
33	64.60	38.48	49.97
34	69.38	39.64	50.04
35	68.24	39.46	50.04
36	67.28	39.18	49.92
37	65.95	40.08	50.05
38	66.87	39.91	50.05
39	63.89	40.34	50.04
40	69.03	39.46	49.88
41	66.70	39.18	49.99
42	69.89	36.52	49.93
43	63.84	40.66	49.94
44	65.16	37.50	49.94
45	64.09	40.42	50.05
46	66.82	40.08	50.01
47	65.35	39.91	50.06
48	66.65	39.73	50.06

Maximum (dB	50.06
Minimum (dB)	49.88
Percentage Difference (%)	0.36
Mean (%)	49.99
Standard Deviation	0.05

Appendix 18

The arrangement of sentences presented in the four speech conditions in each of the stimulus lists from the two groups of stimuli.The sentence numbers are in parentheses.

Speaker		Speaker	Group	One		Stimulus Lists			
						Set	1A	Set	1B
Group	Sex	Number	T	F		T	F	T	F
NC	m	1	12	34		Q-N (12)	L+N (34)	Q+N (12)	L-N (34)
NC	f	1	3	29		Q+N (3)	L-N (29)	Q-N (3)	L+N (29)
NC	m	2	23	37		L-N (23)	Q+N (37)	L+N (23)	Q-N (37)
NC	f	2	8	39		L+N (8)	Q-N (39)	L-N (8)	Q+N (39)
NC	m	3	17	33		Q-N (17)	L+N (33)	Q+N (17)	L-N (33)
NC	f	3	14	38		Q+N (14)	L-N (38)	Q-N (14)	L+N (38)
NC	m	4	9	26		L-N (9)	Q+N (26)	L+N (9)	Q-N (26)
NC	f	4	18	27		L+N (18)	Q-N (27)	L-N (18)	Q+N (27)
NC	m	5	15	36		Q-N (15)	L+N (36)	Q+N (15)	L-N (36)
NC	f	5	1	45		Q+N (1)	L-N (45)	Q-N (1)	L+N (45)
NC	m	6	21	28		L-N (21)	Q+N (28)	L+N (21)	Q-N (28)
NC	f	6	22	35	(—>)	L+N (22)	Q-N (35)	L-N (22)	Q+N (35)
NE	m	1	7	25		Q-N (7)	L+N (25)	Q+N (7)	L-N (25)
NE	f	1	19	40		Q+N (19)	L-N (40)	Q-N (19)	L+N (40)
NE	m	2	11	32		L-N (11)	Q+N (32)	L+N (11)	Q-N (32)
NE	f	2	6	48		L+N (6)	Q-N (48)	L-N (6)	Q+N (48)
NE	m	3	13	41		Q-N (13)	L+N (41)	Q+N (13)	L-N (41)
NE	f	3	10	30		Q+N (10)	L-N (30)	Q-N (10)	L+N (30)
NE	m	4	20	31		L-N (20)	Q+N (31)	L+N (20)	Q-N (31)
NE	f	4	4	44		L+N (4)	Q-N (44)	L-N (4)	Q+N (44)
NE	m	5	2	47		Q-N (2)	L+N (47)	Q+N (2)	L-N (47)
NE	f	5	5	46		Q+N (5)	L-N (46)	Q-N (5)	L+N (46)
NE	m	6	24	42		L-N (24)	Q+N (42)	L+N (24)	Q-N (42)
NE	f	6	16	43		L+N (16)	Q-N (43)	L-N (16)	Q+N (43)

Speaker		Speaker	Group	Two		Set	2A	Set	2B
Group	Sex	Number	T	F		T	F	T	F
NC	m	1	21	29		L+N (21)	Q-N (29)	L-N (21)	Q+N (29)
NC	f	1	22	25		L-N (22)	Q+N (25)	L+N (22)	Q-N (25)
NC	m	2	19	36		Q+N (19)	L-N (36)	Q-N (19)	L+N (36)
NC	f	2	2	28		Q-N (2)	L+N (28)	Q+N (2)	L-N (28)
NC	m	3	4	40		L+N (4)	Q-N (40)	L-N (4)	Q+N (40)
NC	f	3	8	33		L-N (8)	Q+N (33)	L+N (8)	Q-N (33)
NC	m	4	17	37		Q+N (17)	L-N (37)	Q-N (17)	L+N (37)
NC	f	4	14	30		Q-N (14)	L+N (30)	Q+N (14)	L-N (30)
NC	m	5	23	32		L+N (23)	Q-N (32)	L-N (23)	Q+N (32)
NC	f	5	18	35		L-N (18)	Q+N (35)	L+N (18)	Q-N (35)
NC	m	6	13	42		Q+N (13)	L-N (42)	Q-N (13)	L+N (42)
NC	f	6	6	45	(<---)	Q-N (6)	L+N (45)	Q+N (6)	L-N (45)
NE	m	1	16	46		L+N (16)	Q-N (46)	L-N (16)	Q+N (46)
NE	f	1	1	26		L-N (1)	Q+N (26)	L+N (1)	Q-N (26)
NE	m	2	5	27		Q+N (5)	L-N (27)	Q-N (5)	L+N (27)
NE	f	2	9	44		Q-N (9)	L+N (44)	Q+N (9)	L-N (44)
NE	m	3	15	47		L+N (15)	Q-N (47)	L-N (15)	Q+N (47)
NE	f	3	20	48		L-N (20)	Q+N (48)	L+N (20)	Q-N (48)
NE	m	4	24	39		Q+N (24)	L-N (39)	Q-N (24)	L+N (39)
NE	f	4	10	41		Q-N (10)	L+N (41)	Q+N (10)	L-N (41)
NE	m	5	12	34		L+N (12)	Q-N (34)	L-N (12)	Q+N (34)
NE	f	5	3	31		L-N (3)	Q+N (31)	L+N (3)	Q-N (31)
NE	m	6	11	43		Q+N (11)	L-N (43)	Q-N (11)	L+N (43)
NE	f	6	7	38		Q-N (7)	L+N (38)	Q+N (7)	L-N (38)

Shaded cells are stimuli mixed wth noise

Appendix 19

Transcription Record (Set_____)

Participant Code________________ **Date**__________

Procedures:

(1) You are going to listen to **48** statements (some with background noise) through the computer. Each statement is presented only once. After the presentation of each statement, please try as much as you can to write it out **clearly** in standard English spelling in the space provided below. **If you miss any word(s) in the sentence, please draw a dash line.**

(2) After writing down the statement, use the mouse to click on the computer screen one of the three buttons, **"True"**, **"False"**, or **"Unknown"** (in case you are uncertain of the truth value of the statement or you cannot clearly hear the sentence). You will hear a "click" sound that indicates your choice has been registered by the computer. **PLEASE DO NOT CLICK THE BUTTON UNTIL YOU HAVE WRITTEN DOWN THE STATEMENT.**

(3) There is no limitation on time for step (1) and (2). Once you have clicked the button on the screen, the next statement will be presented to you in a short moment.

==

1.____

____

2.____

____

3.____

____

Turn to the next page before you press the button

4. ____________________

____

5. ____________________

____

6. ____________________

____

7. ____________________

____

8. ____________________

____

9. ____________________

____

10. ____________________

____

11. ____________________

____

12. ____________________

____

Turn to the next page before you press the button

13. ____

____

14. ____

____

15. ____

____

16. ____

____

17. ____

____

18. ____

____

19. ____

____

20. ____

____

21. ____

____

Turn to the next page before you press the button

22.

23.

24.

25.

26.

27.

28.

29.

30.

Turn to the next page before you press the button

31.

32.

33.

34.

35.

36.

37.

38.

39.

Turn to the next page before you press the button

40.

41.

42.

43.

44.

45.

46.

47.

48.

Appendix 20

Practice Sentences

(A) True sentences:

(1) You can start a fire with a match.

(2) Most teenagers like rock and roll.

(3) December has thirty-one days.

(4) Grass is green in color.

(B) False sentences:

(5) You can start a fire with a watch.

(6) March has thirty-eight days.

(7) You can buy beer at church.

(8) People play football with a violin.

REFERENCES

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38(4)*, 561-613.

Anglade, Y., & Junqua, J-C. (1990). Acoustic-phonetic study of Lombard speech in the case of isolated-words. *STL Research Reports, 2*, 129-135.

Applebaum, T, Hanson, B., & Morin, P. (1996). Recognition strategies for Lombard speech. *STL Research Reports, 5*, 69-75.

Bauer, R., & Benedict, P. (1997). *Modern Cantonese phonology.* New York: Mouton de Gruyter.

Black, J. (1954). Systematic research in experimental phonetics: 2. Signal reception: Intelligibility and sidetone. *Journal of Speech and Hearing Disorders, 19(2)*, 140-146.

Bond, Z., Moore, T., & Gable, B. (1989). Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *Journal of the Acoustical Society of America, 85(2)*, 907-912.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America, 92(1)*, 593-596.

Cooper, W., Eady, S., & Mueller, P. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America, 77(6)*, 2142-2156.

Cooper, W., & Sorensen, J. (1981). *Fundamental frequency in sentence production.* New York: Springer-Verlag.

Derwing, T., & Munro, M. (1997). Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*, 1-16.

Derwing, T., Munro, M. & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48(3)*, 393-410.

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

Egan, J. (1971). The Lombard reflex: Historical perspective. *Archives of otolaryngology, 94,* 310-312.

Fairbanks, G. (1954). Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders, 19(2),* 133-139.

Fonagy, I., & Fonagy, J. (1966). Sound pressure level and duration. *Phonetica, 15,* 14-21.

Fricke, J. (1970). Syllabic duration and the Lombard effect. *International Audiology, 9,* 53-57.

Fry, D. (1968). Prosodic phenomena. In B. Malmberg (Ed.), *Manual of phonetics* (pp. 365-410). Amsterdam: North-Holland Publishing Company.

Fry, D. (1992). *The physics of speech.* Cambridge, UK: Cambridge University Press.

Gardner, M. (1964). Effect of noise on listening levels in conference telephony. *Journal of the Acoustical Society of America, 36,* 2354-2362.

Guion, S., Flege, J., Liu, S., & Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21(2),* 205-228.

Haagen, C. (1946). *Intelligibility measurements.* Speech Monographs, 13(2) 4-7.

Haimson, B., & Elfenbein, M. (1985). *Experimental methods in psychology.* New York: McGraw-Hill, Inc.

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

Hatch, E., & Lazaraton, A. (1991). *The research manual.* New York: Newbury House.

Hazan, V., & Simpson, A. (1998). The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Speech Communication, 24,* 211-226.

He, S. (1987). Several aspects in the comparison between Chinese and English in the teaching of English phonology. *Foreign Languages, 52,* 70-73.

Junqua, J-C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America, 93(1),* 510-524.

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

Junqua, J-C., & Anglade, Y. (1990). Acoustic and perceptual studies of Lombard speech: Application to isolated-words automatic speech recognition. *STL Research Reports, 2,* 73-81.

Kehoe, M., Stoel-Gammon, C., & Buder, E. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research, 38,* 338-350.

Keller, E. (1994). *Signalyze user's manual.* Lausanne, Switzerland: InfoSignal Inc.

Korn, T. (1954). Effects of psychological feedback on conversational noise reduction in rooms. *Journal of the Acoustical Society of America, 26,* 793-794.

Kryter, K. (1946). Effects of ear protective devices on the intelligibility of speech in noise. *Journal of the Acoustical Society of America, 18,* 413-417.

Ladefoged, P. (1993). *A course in Phonetics.* Fort Worth, TX: Harcourt Brace Jovanovich, Inc.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35(4),* 451-453.

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

Leder, S., & Spitzer, J. (1993). Speaking fundamental frequency, intensity, and rate of adventitiously profoundly hearing-impaired adult women. *Journal of the Acoustical Society of America, 93(4)*, 2146-2151.

Li, C-N. (1999). The effect of the Lombard reflex on speech produced by Cantonese speakers of English. Poster session presented at the 138th Acoustical Society of America meeting (November 1-5, 1999), Columbus, OH.

Lombard, E. (1911). Le Signe de l'Elévation de la Voix. *Ann. Maladiers Oreille, Larynx, Nez, Pharynx, 37*, 101-119.

Loren, C., Colcord, R., & Rastatter, M. (1986). Effects of auditory masking by white noise on variability of fundamental frequency during highly similar productions of spontaneous speech. *Perceptual and Motor Skills, 63*, 1203-1206.

Miller, G., Heise, G., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41(5)*, 329-335.

Munro, M. (1995). Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition, 17*, 17-34.

Munro, M. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition, 20*, 139-154.

Munro, M., & Derwing, T. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing, 11*, 253-266.

Munro, M., & Derwing, T. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45(1)*, 73-97.

Munro, M., & Derwing, T. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38(3),* 289-306.

Munro, M., & Derwing, T. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 48(2),* 159-182.

Munro, M., & Derwing, T. (1999). *The effects of speaking rate on the comprehensibility and accentedness of L2 speech.* Manuscript submitted for publication.

Palmer, F. R. (1981). *Semantics.* Cambridge, UK: Cambridge University Press.

Peterson, G., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32(6),* 693-703.

Picheny, M., Durlach, N., & Braida, L. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96-103.

Picheny, M., Durlach, N., & Braida, L. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434-446.

Pickett, J. (1956). Effects of vocal force on the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 28(5),* 902-905.

Pisoni, D. (1996). Word identification in noise. *Language and Cognitive Processes, 11(6),* 681-687.

Pisoni, D., & Dedina, M. (1986). Comprehension of digitally encoded natural speech using a sentence verification task (SVT): A first report. *Research on speech perception (No. 12).* Bloomington, IN: Indiana University.

Pollack, I., & Pickett, J. (1958). Masking of speech by noise at high sound levels. *Journal of the Acoustical Society of America, 30(2),* 127-130.

Quach, W. (1998). *Linguistic equivalence of the sentence format of the Test of Children's Speech.* Unpublished MSLP thesis, University of Alberta.

Reissland, N., & Snow, D. (1996). Maternal pitch height in ordinary and play situations. *Journal of Child Language, 23*, 269-278.

Rivers, C., & Rastatter, M. (1985). The effects of multitalker and masker noise on fundamental frequency variability during spontaneous speech for children and adults. *Journal of Auditory Research, 25*, 37-45.

Roach, P. (1982). On the distinction between stress-timed and syllable-timed languages. In D. Crystal (Ed.), *Linguistics and controversies* (pp. 682-688). London: Edward Arnold.

Saito, M., Fukudome, K., & Tsumura, T. (1999). Intelligibility improvement of cellular phone speech in noisy environment. Poster session presented at the 138th Acoustical Society of America meeting (November 1-5, 1999), Columbus, OH.

Sakiey, E., & Fry, E. (1979). *3000 instant words.* Highland Park, NJ: Dreier Educational Systems.

Steeneken, H., & Hansen, J. (1999). Speech under stress conditions: Overview of the effect on speech production and on system performance. *ICASSP, 4*, 2079-2082.

Van Summers, W., Pisoni, D., Bernacki, R., Pedlow, R., & Stokes M. (1988). Effects of noise on speech production: acoustic and perceptual analyses. *Journal of the Acoustical Society of America, 84(3)*, 917-928.

Webster, J., & Klumpp, R. (1962). Effects of ambient noise and nearby talkers on a face-to-face communication task. *Journal of the Acoustical Society of America, 34(7)*, 936-941.

Yamazawa, H., & Hollien, H. (1992). Speaking fundamental frequency patterns of Japanese women. *Phonetica, 49*, 128-140.