Fundamental frequency of English words in sentences produced by nonnative female speakers in noise [1]

Chi-Nin Li
Simon Fraser University

This study examined the effect of noise on fundamental frequency (Fo) values of English words in sentences spoken by native Hong Kong Cantonese speakers and native English speakers. Five female speakers from each of the two groups read a list of English declarative sentences, once in quiet and once in a background of masking noise at 70 dB(A). The mean Fo values of three key words in each sentence produced in the two speaking conditions were measured. The mean Fo values were significantly higher in speech produced in noisy conditions than in quiet conditions. The nonnative speakers exhibited higher Fo values than did the native English speakers in both conditions. The present study confirmed one of the manifestations of the Lombard reflex that speakers tend to increase fundamental frequencies in noisy conditions. The findings suggest that like female native English speakers, female Cantonese speakers of English exhibit higher fundamental frequency when speaking in noisy environments.

1 Introduction

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). When the noise stops, the voice intensity returns to a normal level (Egan, 1971; Junqua, 1996; Lane & Tranel, 1971). This phenomenon was first described in 1911 by Etienne Lombard, the first oto-rhino-laryngologist in France (Lane & Tranel, 1971).

Lane and Tranel (1971) and, more recently, Junqua (1996), have conducted detailed surveys of the literature on the Lombard reflex 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 fundamental frequency, voice level, 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, speaking rate may be reduced when speech is produced in a noisy environment (Hanley & Steer, 1949). 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.

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.

[1] The research reported here is part of a larger project. I would like to thank Dr. Murray Munro (SFU) for his advice on the original study, and the audience at NWLC 2001 (UBC) for their comments. The study was supported through a grant by SSHRC to Dr. Munro.

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.

Among all the features associated with speech produced in noise, a change in voice level is the most obvious manifestation of the Lombard effect. Researchers agree that when speakers talk in a background of noise, their speech becomes more intense (Bond et al., 1989; Lane & Tranel, 1971; Steeneken & Hansen, 1999). 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 and in multitalker noise were higher than in quiet. As was the case for the children, the adult participants utilized mean Fo values that were lower in quiet than in either white noise or multitalker noise. 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, 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 than in the quiet condition.

As noted by Lane and Tranel (1971), the Lombard reflex has been considered an important phenomenon in the analysis of speech communication in noise. Because research associated with the Lombard effect has been conducted mainly with native speakers, little is known about its impact on foreign-accented speech. In this study, an attempt is made to examine the production of Lombard speech by Hong Kong Cantonese speakers of English. Fundamental frequencies of English words in simple declarative sentences produced under normal and noisy conditions by native Cantonese and native English speakers are examined. Findings from previous studies of the Lombard effect indicate that in a noisy environment, 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 utilize higher fundamental frequency values in noisy than in quiet conditions.

In addition, results from studies of second language speech show that Mandarin speakers of English exhibit higher fundamental frequency values than do their native counterparts (Munro, 1995). Since both Mandarin and Cantonese are Chinese languages, the Cantonese speakers, like the Mandarin speakers, are expected to exhibit higher fundamental frequency values than those of the native speakers of English. The results will be interpreted in terms of characteristics of the native languages of the speakers and of the Lombard reflex.

2 Method

Five female native Hong Kong Cantonese speakers served as subjects[2]. 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 21.6 years at the time of study with a range of 20 to 26 years. They had lived in Canada for a mean of 25 months with a range of nine to 45 months. They were undergraduate students at Simon Fraser University. 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.

Five female native speakers of English 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 24 years with a range from 21 to 27 years. All speakers 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.

A list of 10 statements was used as stimulus (see Appendix). Similar sets of statements have been used in other earlier studies (Derwing, Munro, & Wiebe, 1998; Munro, 1998; Munro & Derwing, 1995, 1999). Each item was a single clause sentence of five to eight words. All vocabulary items in the statements were listed as high frequency words by Sakiey and Fry (1979).

Individual recordings were made in a sound-treated room in the Phonetics Laboratory of Simon Fraser University. 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. Masking noise at 70 dB(A) was presented to the speakers through the headphones using the JVC TD-W709 Cassette Deck. 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.

The participants read the stimulus sentences on the test lists under two conditions: quiet and noise. 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. A 5-minute break was provided between recordings in the two conditions.

Prior to the recordings, the participants 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 recording levels were adjusted as appropriate. During recording, 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.

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. Measurements of the mean Fo

[2] In the original study (Li, 2000), 12 speakers (6 male, 6 female) from each of the two groups participated in the experiment. A set of 48 simple true and false English statements was used as stimuli.

values of each of the three key words in each of the 10 stimulus sentences produced in quiet and noisy conditions were made with the PitchEditor in Praat (version 3.9.12) using the autocorrelation algorithm. The analysis filter was set from 135- to 535-Hz for the female voices at every 5 msec (cf. Loren et al., 1986; Rivers & Rastatter, 1985).

3 Results and discussion

The mean Fo values of each of the three key words in each of the 10 sentences were measured, and the overall mean value for each speaker was computed. The mean Fo values were submitted to a mixed-design ANOVA with Native Language of Speakers (NL), Cantonese and English, as a between-subjects factor, and Speaking Conditions (SC), Quiet and Noise, as a within-subjects factor. The analysis revealed a significant effect of the NL, F (1, 8) = 5.41, $p < 0.05$, indicating that the mean Fo values for Cantonese speakers in Quiet (M = 225.27 Hz) and in Noise (M = 253.67 Hz) were significantly higher than those of English speakers in Quiet (M = 201.48 Hz) and in Noise (M = 230.26 Hz). The analysis also revealed a significant effect of the SC, F (1, 8) = 120.81, $p < 0.0001$. All speakers exhibited mean Fo values that were higher in the Noise condition than in the Quiet condition. However, no significant interaction was found for NL x SC, F (1, 8) = 0.005, $p >$ 0.05. Figure 1 illustrates the mean Fo values with respect to the native language and the two speaking conditions.

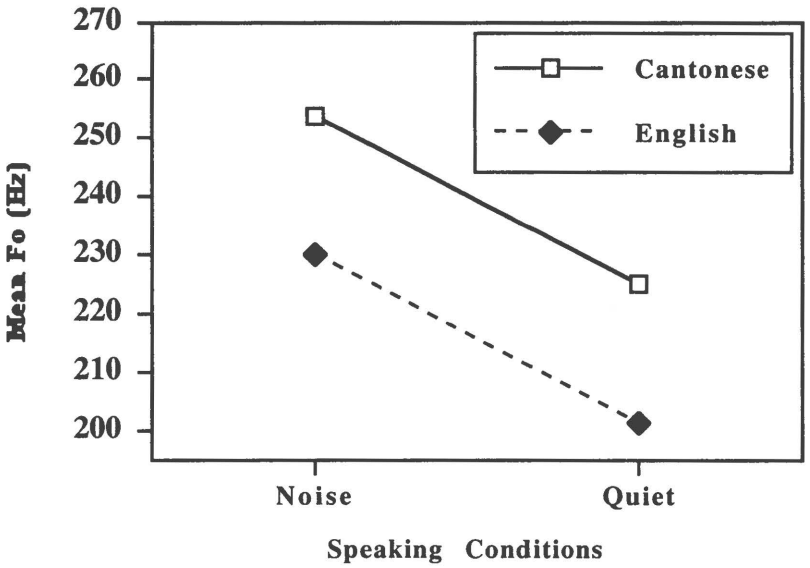

Figure 1. Mean fundamental frequency value (Hz) according to the native language under the two speaking conditions.

First of all, the findings confirmed one of the manifestations of the Lombard effect that speakers generally have a tendency to utilize higher fundamental frequency in noisy than in quiet conditions. The results are consistent with earlier research findings (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. Lane and Tranel (1971) reported 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 addition, 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. The reasons for the Cantonese speakers' utilizing an overall higher fundamental frequency than the English speakers are not fully understood. Yamazawa and Hollien (1992) suggested that 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. Yamazawa and Hollien (1992) found that female Japanese speakers of English exhibited higher speaking fundamental frequency in reading English passage than did native English speakers. They suggested that the tone aspect of Japanese, a syllable-timed pitch accent language, may account for the addition of higher frequencies that leads to the higher mean Fo relative to that of the native English speakers. If this is the case, it is possible that some carry-over from Cantonese, a tonal language, for the nonnative speakers influences English productions in such a way that sentential Fo is higher than that of the native English speakers. Nevertheless, this language-dependent phenomenon needs further research.

4 Conclusion

The purpose of this experiment was to explore the effect of the Lombard reflex on sentences spoken by nonnative speakers of English. This was done by examining differences in fundamental frequency in English words 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. All of the speakers exhibited a higher Fo 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 quiet, but also in noisy conditions. It is hoped that 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.

References

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.

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.

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.

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

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.

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.
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.
Li, C-N. 2000. *The role of the Lombard reflex in the production and perception of Cantonese–accented English.* Unpublished master's thesis. Simon Fraser University.
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.
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. 1995. 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. 1999. *The effects of speaking rate on the comprehensibility and accentedness of L2 speech.* Manuscript submitted for publication.
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.
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.
Yamazawa, H. & Hollien, H. 1992. Speaking fundamental frequency patterns of Japanese women. *Phonetica*, 49, 128-140.

Appendix

List of stimulus sentences. Key words used for Fo analysis are bolded.

1. **Bread** is **made** with **flour**
2. An **engine** is a **part** of a **ship**
3. A **tiger** is **bigger** than a **cat**
4. You can **tell time** with a **watch**
5. A **pigeon** is a **kind** of **bird**
6. **Seven** is an **odd number**
7. **Hot** and **cold** are **opposites**
8. **Japan** is a **wealthy country**
9. **Ships travel** on the **water**
10. **Red** and **green** are **colors**