INTELLIGIBILITY OF FOREIGN-ACCENTED LOMBARD SPEECH

Chi-nin Li
Dept. of Linguistics, Simon Fraser University, 8888, University Drive., B.C., Canada, V5A 1S6. clia@sfu.ca

1. INTRODUCTION

The Lombard reflex occurs when normal-hearing speakers modify their vocal effort while speaking under noisy conditions (Lane & Tranel, 1971; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Speech produced in noisy conditions has been reported to be different from speech produced in quiet conditions in a number of ways: increase in vowel duration, a shift in the first two formants, decrease in speaking rate, and increase in fundamental frequency (Junqua, 1996). Also, Lombard speech is more intelligible than normal speech when presented at an equal level of masking noise (Van Summers et al., 1988).

Because research associated with the Lombard reflex has been conducted mainly with native speakers, little is known about its impact on foreign-accented speech. The purpose of this experiment is to examine the effects of masking noise on the intelligibility of Lombard speech produced by non-native English speakers using a sentence-transcription task and a sentence-verification task.

2. METHOD

2.1 Speech material

A list of 24 true and 24 false statements were recorded by 12 Hong Kong Cantonese speakers (6 female, 6 male) and a comparison group of native speakers of Canadian English (6 female, 6 male) in a sound-treated room. All speakers passed a pure-tone hearing screen. They produced in a conversational manner the list of sentences once under each of two speaking conditions: quiet and noise. Under the noise condition, cafeteria–like masking noise of 70 dB was fed through headphones to the speakers. No noise was presented to the speakers in the quiet condition. Stimuli were digitized at a 22.05 kHz sampling rate with 16-bit amplitude quantisation.

Four different sentences (2 true, 2 false) produced in each of the speaking conditions were selected from each of the 24 speakers for a total of 192 utterances. From each speaker, the four specific sentences produced in the quiet condition (quiet) were identical to those produced in the noise condition (Lombard). Selected Lombard and quiet speech samples were mixed with the same type of masking noise used in the production task at a constant signal-to-noise (S/N) ratio (noisy), and were presented to listeners with unmasked sentences (clean). This resulted in four different speech conditions: (1) Clean quiet (Q-N), (2) Noisy quiet (Q+N), (3) Clean Lombard (L-N), and (4) Noisy Lombard (L+N). Four separate stimulus sets were prepared. Each set consisted of the full complement of 48 test items (24 true, 24 false). For each set, individual speakers were represented twice, once producing a true sentence and once producing a false sentence. In addition, each of the stimulus sets was balanced for speakers' native language and gender, and the speech conditions. The root-mean-square (RMS) amplitude for each of the 192 stimulus sentences was equated.

2.2 Listeners

The listeners were eight female native speakers of Canadian English. They were born and raised in Vancouver, British Columbia. Their mean age was 21.5 years, and all passed a pure-tone hearing screen. The listeners in groups of two were randomly assigned to listen to one of the four sets of stimuli.

2.3 Procedure

Stimuli were presented via headphones at a comfortable listening level to the listeners using a custom response-collection program. The listeners transcribed the stimulus sentences in standard orthography. They then verified the truth value of the statement by clicking one of three buttons shown on a computer screen ("True", "False", or "Unknown").

3. RESULTS

3.1 Transcription Scores

Scores were assigned to each sentence by computing the percentage of words that were correctly transcribed. Figure 1 illustrates the mean values for the sentences presented under the four speech conditions. In each of the four speech conditions, the Cantonese speakers' sentences were less correctly transcribed than those of the native productions. A two-way repeated measures ANOVA revealed significant effects of Native Language of Speakers (NL), F (1,7) = 54.99, $p <$ 0.01, and Speech Condition (SC), F (3,21) = 15.06, $p <$ 0.0001. Post hoc analyses revealed that noisy utterances (Q+N and L+N) received significantly lower scores than did clean utterances (Q-N or L-N), $ps < 0.05$, and that listeners' scores in the condition of L+N were significantly higher than those in the condition of Q+N, $ps < 0.05$. No significant difference was found between the conditions of L-N and Q-N, $ps > 0.05$.

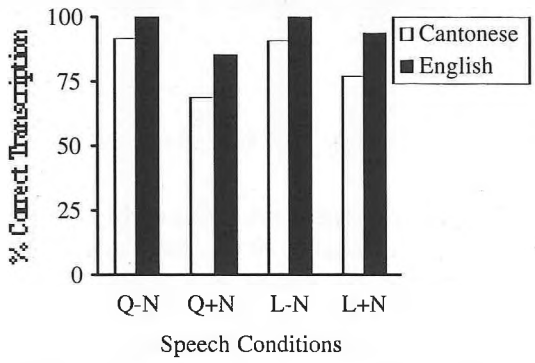

Fig. 1. Percent correct transcription scores for the stimulus sentences presented in the four speech conditions.

3.2 Verification Scores

The verification scores were determined by summing the number of correct true and false responses. The choice of "Unknown" was considered as incorrect. The mean values are given in Figure 2. Cantonese speakers' sentences were correctly verified less frequently than were the native productions in all speech conditions. A repeated measures ANOVA yielded significant 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, $F(3, 21) = 3.24$, $p < 0.05$. For the non-native productions, statistical analyses revealed that clean sentences (Q-N and L-N) were correctly verified more often than were noisy sentences (Q+N and L+N), $ps < 0.05$. Other pairwise comparisons (L-N vs. Q-N and L+N vs. Q+N) were not significant, $ps > 0.05$. For the English speakers, utterances presented in Q+N received significantly lower scores than did those presented in Q-N, L-N or L+N, $ps < 0.05$. Nevertheless, there were no significant differences among the three speech conditions, Q-N, L-N, or L+N, $ps > 0.05$.

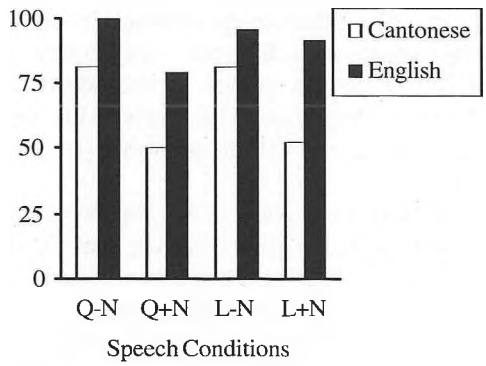

Fig. 2. Percent correct verification scores for the stimulus sentences presented in the four speech conditions.

4. DISCUSSION

Cantonese speakers' utterances were found to be less intelligible than were the native English productions. The sentences spoken by the Cantonese speakers were less correctly transcribed and verified than were the native-produced sentences in each of the four speech conditions.

Unmasked sentences were better perceived than were noisy sentences in most cases. Utterances presented in the 'clean' conditions (Q-N and L-N) were more correctly transcribed than were sentences presented in the 'noisy' conditions (Q+N and L+N), indicating that the masking noise degrade the intelligibility of both the Lombard and quiet speech. This pattern was again observed in the verification scores for Cantonese speakers' sentences, but not for English speakers' productions. For the English speakers, there was no significant difference in the verification scores between their noisy (L+N) and clean (L-N) Lombard speech.

For both groups of speakers, their noisy Lombard speech (L+N) was correctly transcribed more frequently than were the noisy quiet sentences (Q+N). The improvement in intelligibility is likely due in part to the characteristics of the Lombard speech. For the verification task, such a pattern was found for the native English speakers, but not for the Cantonese speakers. No significant difference in verification scores was found between the non-native Lombard speech and quiet speech when presented in noisy conditions. The discrepancy in patterns for the intelligibility of the non-native noisy sentences may be due to the different nature of the two tasks involved, as it has been suggested that the verification measure appears to be a coarse evaluation of intelligibility (Pisoni & Dedina, 1986).

REFERENCES

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.

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.

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.

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.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Murray Munro for his advice on this research.