PERCEPTION OF ENGLISH /r/ AND /l/ SPEECH CONTRASTS BY NATIVE KOREAN LISTENERS WITH EXTENSIVE ENGLISH-LANGUAGE EXPERIENCE

D.G. Jamieson and K. Yu

Hearing Health Care Research Unit, The University of Western Ontario, London, ON, Canada N6G 1H1

1. INTRODUCTION

Native speakers of one language often have difficulty perceiving and producing the sounds of other languages correctly. For example, native speakers of Korean often have difficulty perceiving the English /r/-/l/ sounds [3,12,17].

While English has two separate labels for the sounds we call /r/ and /l/, the Korean language groups these sounds into a single category [3]. In Korean, the sound heard by English-speaking listeners as /r/ occurs only intervocalically; that heard as /l/ occurs only in word-final position [13]. Particular problems may arise when communicating against a background of noise. Non-native listeners may display greater difficulty than native listeners under difficult listening conditions [16,20].

Notwithstanding these factors, it is now clear that laboratory training can quickly and substantially improve listeners' perceptions of non-native speech contrasts [9,10,14,21]. In the case of /r/-/l/ identification, Logan et al. [14] demonstrated that training with natural English tokens improved /r/-/l/ identification for native speakers of Japanese.

In view of the foregoing, we sought to examine the degree to which very extensive exposure to the English language as an adult, through many years of living in a predominantly English-language environment, affects the perception of "difficult" English speech contrasts such as /r/-/l/. The present study was therefore designed to study aspects of the speech perception abilities of adults who were native speakers of Korean but who had lived for many years in an English-speaking environment.

2. EXPERIMENTAL STUDIES

2.1 Method

Subjects. Thirty native speakers of Korean (17 male and 13 female), aged 35 to 58 ($\underline{M}$ = 49 years) volunteered to participate in this experiment. All were native to Korea but had resided in Canada for between 10 and 28 years ($\underline{M}$ = 20.33 years) immigrating between the ages of 20 to 40 ($\underline{M}$ = 29 years). None reported any history of speech or hearing difficulties. All subjects had studied English (mean time = 7.13 years) before coming to Canada. Twenty subjects had also studied in ESL programs in Canada (mean time = 6.27 months) with the emphasis on speaking/conversational/ pronunciation skills. Sixteen of the subjects had a high school level of education, 14 had a university education. Subjects communicated in English 10 to 80 percent of the time ($\underline{M}$ = 44%) and rated their abilities to read and understand spoken English as being "fair". Six native Canadian English speakers with normal hearing also participated.

Stimuli and Equipment. Our test battery consisted of three tests of speech perception. The adaptive SRT (ASRT; [4]) provides an efficient, accurate, and reliable estimate of a listener's speech reception threshold for spondees. The SRT was administered the in quiet and in a background of white noise.

The UWODFD [5] is a test of speech intelligibility, standardized for central Canadian English. It consists of 21 nonsense syllable stimuli; the target is the middle consonant of a VCVC word. All consonants are presented in the same context (A_IL). During each trial, one stimulus was presented, and the subject indicated which word was heard by selecting one of 21 response alternatives displayed on a video monitor. The UWODFD was administered in quiet and in noise at 4 Signal to Noise Ratios (SNRs; +10, +5, 0, and -5 dB). Listening conditions were presented in sequence from the most favorable (i.e., in Quiet) to the most difficult (i.e., at -5 dB SNR), with the order of stimulus presentation randomized within each level.

The ability to distinguish English /r/ versus /l/ contrasts was measured using a two alternative forced-choice [r-l] identification task with a subset of the stimuli used by Logan et al. [14]. The original 12-bit, 10kHz sampled signals used in the Logan et al. [14] study, were processed digitally to convert them to the format required for our computer programs. The resulting signals were stored on computer disk as 16-bit, 16kHz signals in CSRE format [11].

This [r-l] test consisted of five minimal pairs (e.g., rock-lock) within each of five phonetic environments: initial singleton (IS), initial consonant cluster (IC), medial (M), final consonant cluster (FC), and final singleton (FS). Five native English speakers (3 male and 2 female) produced each of the ten words, contrasting /r/ and /l/ in the phonetic environments mentioned. Each of the five phonetic environments was tested in a separate session containing five blocks of trials -- each with a different talker. Immediately after a stimulus was presented, subjects using a computer mouse to indicate whether they had heard an /l/ or an /r/ sound by selecting one of two corresponding response alternatives, "R" or "L", displayed on a video monitor. No feedback was provided. Each of the five experimental tests was administered in 5 listening conditions: (1) in quiet and (2) in noise at 4 SNRs (10, 5, 0, and -5 dB). Listening conditions were presented from least to most difficult for each of the phonetic environments, with both the block order and stimuli, randomized.

Throughout the study, subjects were tested individually, while seated with the monitor and keypad, within an Industrial Acoustics Com-

pany (IAC) double-walled acoustically-shielded room. All aspects of stimulus sequencing and presentation (in both noise and quiet conditions), response recording, and experimental control were carried out using the experiment generator and controller utility contained in the CSRE 4.5 software package [6]. During all aspects of testing and training, signals were replayed over an Ariel DSP-56 digital to analog converter, attenuated and amplified (AMCRON D-75) to achieve the required signal level, and presented to listeners through Etymotic Research ER-2 insert phones.

3. RESULTS

Our analyses addressed four questions: (1) how accurate were the native Korean listeners in identifying English words under optimal listening conditions? (2) was the performance of native Korean listeners differentially affected when the listening conditions became more difficult because of background noise? (3) in what way did /r/-/l/ identification accuracy depend on syllable position for the native Korean listeners? and (4) how were identification errors distributed across target sounds and responses?

Accuracy Under Optimal Listening Conditions. The results from the three tests, in quiet, show that the overall level of accuracy under optimal listening conditions was high for both native Korean listeners and for listeners who were native speakers of Canadian English. Figure 1 shows that listeners whose native language was Korean were able to achieve 70% correct identification of spondaic words at a signal level attenuation level averaging 59 dB, while listeners who were native speakers of Canadian English achieved the same level of performance with 64 dB of attenuation. This result indicates that native speakers of Korean required a signal level which was ~5 dB higher than native speakers of English, in order to achieve the same level of accuracy.

Figures 2 and 3 show that native speakers of Korean achieved 76 and 71% correct overall, in the UWODFD and [r-l] tests, respectively, under quiet listening conditions. Native English speakers achieved 94 to 98% correct in the UWODFD and [r-l] tests, respectively, under the same conditions.

Accuracy Under More Difficult Listening Conditions. Under less optimal listening conditions, the performance of the Korean listeners declined rapidly. Figure 1 shows that listeners whose native language was Korean could tolerate a signal attenuation level averaging 28.5 dB when identifying spondaic words in a background of noise Listeners who were native speakers of Canadian English achieved the same level of performance at an attenuation of 31.5 dB SNR. Thus, native Korean listeners required listening conditions that were 3 dB more favorable, on average, to achieve a level of performance that was comparable to that of native English speakers. Thus, for the ASRT task, there was no indication that performance declined more rapidly in a noise background for our native Korean subjects.

For native speakers of Korean, performance declined when listening in noise, falling .94%/ per dB for the [r-l] test to just 50.5% (chance = 50%) in the -5 dB SNR condition, and falling 3.1% per dB for the UWODFD test, to just 14.9% (chance ~ 5%) in the -5 dB SNR condition. Performance also declined with both tests for native speakers of Canadian English, but this decline was more precipitous,

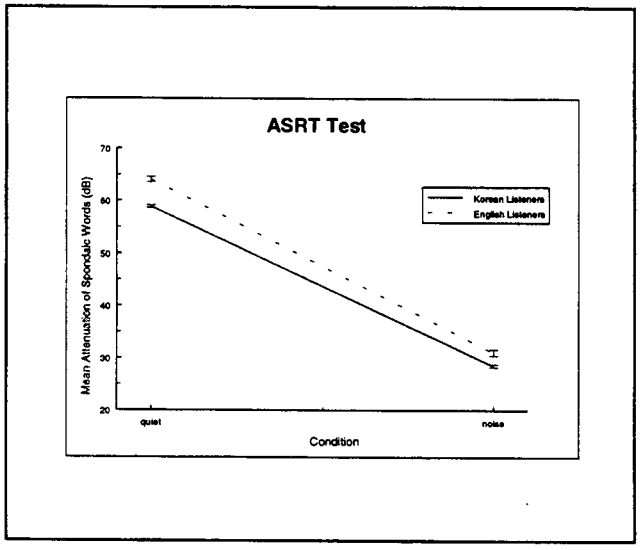

Figure 1. Speech Reception Threshold (SRT) for spondaic words, measured using the ASRT test, in Quiet and in the presence of background noise. Data are for 30 listeners who were native speakers of Korean, and for six listeners who were native speakers of Canadian English. The attenuation level reflects the amount by which the signal could be attenuated while still achieving 70.7% correct identification of target words: superior performance is thus indicated by greater attenuation levels.

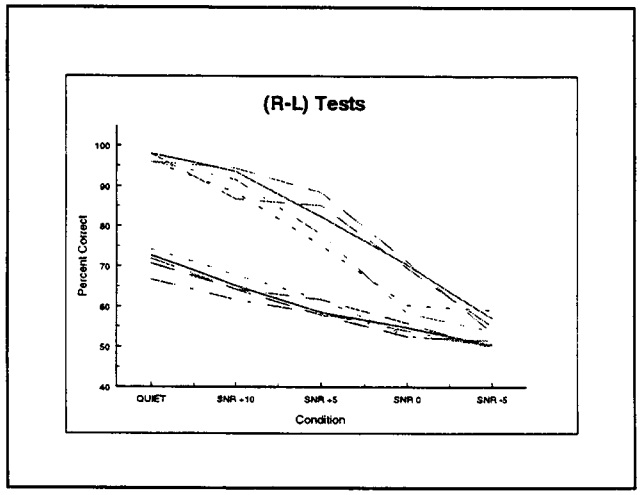

Figure 2. Identification of English /r/ and /l/, by listeners who were native speakers of Korean, and by listeners who were native speakers of Canadian English, as a function of the ratio of the signal level to the level of background noise. The upper fan of curves is for native English-speaking listeners for target sounds in the various syllable positions, while the lower fan of curves presents the comparable data for native Korean-speaking listeners.

reflecting the higher initial levels of performance in quiet: the performance decrement averaged 2.3% per dB with the [r-l] test and 4.1% per dB with the UWODFD test.

Accuracy as a Function of the Position of the Target in the Syllable. Figure 1 shows that the accuracy with which /r/ and /l/ were identified depended to a considerable extent on syllable position. Overall accuracy was highest in the Final Singleton (FS) position and lowest in the Initial Consonant Cluster (ICC) position for Korean listeners, with the difference between these positions being 7.6% in quiet, 6.1% at +10 dB SNR, and 3.6% at +5 SNR in the [r-l] test. It is perhaps noteworthy that native English speakers showed a 6.8% difference in accuracy between these two positions at the +10 SNR level, but there was no apparent difference at other levels.

These generalizations are limited, however, by an interaction between target sound and syllable position. When the target sound was /l/, identification accuracy was essentially equivalent i initial singleton, initial consonant cluster, medial and final cons cluster positions (~72 to 74% correct in quiet for example), but fell to below 69% correct in final singleton position. Performance in the final singleton position was also ~5% lower than in other positions in the +10 dB and -5 dB conditions, and it was ~10% lower than in other positions in the +5 and 0 dB conditions. This pattern of difficulty, even when performance was at near chance levels, suggests that there was a modest bias to respond "R", when the target signal occurred in final singleton position.

When the target sound was /r/, performance was best with the target in the final singleton position (~80% in quiet) and in the initial singleton positions (~74% in quiet). Identification accuracy fell to 67-69% correct in medial and final consonant cluster positions in quiet and to just 59% in initial consonant cluster position. Performance continued to be best for /r/ targets in final singleton position, at all signal levels, with the difference growing to ~10 % relative to the initial singleton position in the +10 and +5 dB conditions, and remaining at ~5% even in the -5 dB condition, when overall performance was essentially at chance. This pattern of results likely reflects the combination of a bias to respond "R" for targets in syllable-final position, possibly in combination with increased salience of cues to the /r/-/l/ distinction for targets occurring in that position.

Distribution of Identification Errors. Examination of Figure 3 shows that when the format of the task does not constrain subjects' responses to just "L" and "R", the observed perceptual confusions are by no means limited to confusing /r/ with /l/ and vice versa. In fact, when listening in quiet during the UWODFD test, 59% of the confusion errors Korean listeners made when /r/ targets were presented were "W" responses, compared to 34% "L" responses and 7% "Y" responses. Under the same conditions, 59% of confusion errors when /l/ targets were presented were "R" responses, while 37% were "N" responses and 4% were "W" responses. With the UWODFD test, in the +10 dB condition, "L" was never provided as a confusion response to /r/, with "W" being given 74% of the time and "Y" 26% of the time. Under the same conditions, when the target was /l/, "N" was the most common response (34%), followed by "W" (32%), "R" (26%) and "Y" (8%).

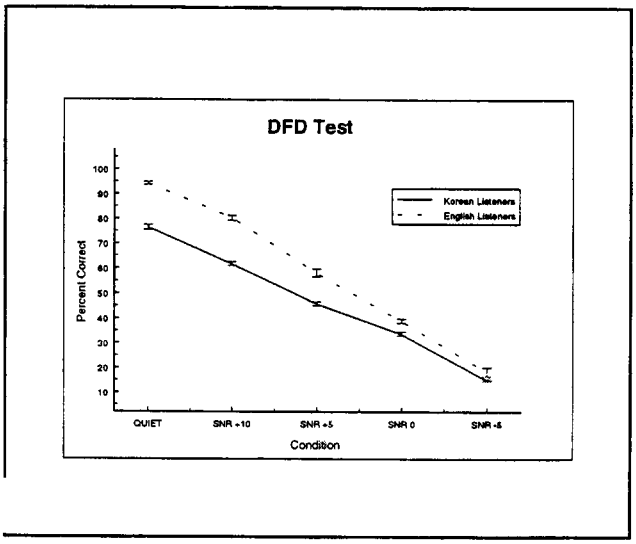

Figure 3. Identification of English syllable-medial consonants in the UWODFD test, by listeners who were native speakers of Korean (lower curve), and by listeners who were native speakers of Canadian English (upper curve), as a function of the ratio of the signal level to the level of background noise.

4. GENERAL DISCUSSION

The present study provides clear evidence that native speakers of the Korean language have considerable difficulty identifying tokens of the English language /l/ and /r/ sounds. The degree of difficulty encountered varies as a function of the phonetic environment in which the target sounds appear, and with the listening conditions.

ASRT test results indicate that even in a simple closed-set test, with very little stimulus uncertainty under favorable (quiet) listening conditions, native speakers of Korean require a signal level that is 5 dB more favorable than native speakers of English, to achieve good speech intelligibility. The noise background affected Korean and native English-speaking listeners similarly, but reduced the advantage for native English listeners to approximately 3 dB.

These results confirm that non-native listeners do operate at a moderate disadvantage to native listeners even after many years of intensive experience with the English language. However, the disadvantage is seen most clearly under favorable listening conditions, not under degraded conditions. This result was observed under test conditions that were designed specifically to preclude the use of higher-level grammatical and other cognitive/linguistic knowledge. As one might expect that native speakers would have superior cognitive/linguistic skills, the advantage for native listeners' would likely be still greater in spoken language comprehension tasks where such knowledge could be applied.

The [r-l] test confirmed a substantial advantage for native English-speaking listeners in the identification of /r/ and /l/. As with the ASRT, this advantage was greatest under the most favorable listening conditions (in quiet and at +10 dB SNR), declining to nil at the

least favorable SNR (-5 dB). This result is therefore consistent with that for the ASRT test, in that the greatest advantage conveyed by native English-language experience is seen in the most favorable listening conditions.

As with studies with Japanese listeners [15,18], our listeners identified English /r/ and /l/ sounds most accurately when these target sounds were presented in final singleton position. The liquid that occurs in word-final position in the Korean language is /l/-like for native speakers of English, but neither /r/ nor /l/ occur in this position in Japanese. This effect is therefore not likely to reflect aspects of any relative similarities between the phonemic categories of the new language and those of the listener's native language. Rather, the good performance may well reflect a combination of the relative salience of cues intrinsic to the /r/ and /l/ sounds which are available in final singleton position (overall, identification accuracy was highest for targets in that position), in combination with a bias, by our Korean listeners, in favor of identifying any "ambiguous" syllable-final token as "R", rather than as "L".

The UWODFD test further confirmed the results of with the ASRT and [r-l] tests, showing the greatest advantage for native English listeners under the most favorable listening conditions. However, examination of the details of the confusion responses of Korean listeners shows that the difficulties in correctly determining that /r/ or /l/ has been presented, does not imply that /r/ is most commonly misheard as "L" and that /l/ is most commonly misheard as "R". In fact, the most common error responses to /r/ presentations were "W" responses, while 37% of error responses to /l/ presentations were "N". These results suggest that the two-alternative forced choice paradigm fails to characterize /r/-/l/ perception for non-native listeners. Future work may profit from the use of a wider range of response alternatives when assessing speech perception as well as for training listeners to perceive non-native contrasts.

5. REFERENCES

1. Best, CT, McRoberts, GW and Sithole, NM "Examination of perceptual reorganization for non-native speech contrasts: Zulu click discriminations by English-speaking adults and infants", *J Exper Psych: Human Percep and Perform, 14*(3), 345-60, 1988.
2. Ocke-Schwen Bohn, "What determines the perceptual difficulty encountered in the acquisition of non-native contrasts?", in K Elenius and P Branderud (Eds.) *Proceeds Internat Congress of Phonetic Sci, 4,* Stockholm: KTH and Stockholm University: Stockholm, 84-91, 1995.
3. Borden, G, Gerber, A and Milsark, G "Production and perception of the /r/-/l/ contrast in Korean adults learning English", *Language Learning, 33*(4), 499-526, 1983.
4. Cheesman, MF "An automated technique for estimated speech-reception thresholds in multi-talker babble" *J Speech-Lang Path and Aud, 16*, 223-7, 1992 .
5. Cheesman, MF and Jamieson, DG "Development, evaluation and scoring of a nonsense word test suitable for use with speakers of Canadian English", *Canadian Acoustics, 24,* 3-11, 1996.
6. *Computerized Speech Research Environment (CSRE) Version 4.5 (1996)*, Avaaz Innovations Inc., London, Canada, 1996.
7. Goto, H "Auditory perception by normal Japanese adults of the sounds "L" and "R"", *Neuropsychologia, 9*, 317-23, 1971.
8. Henly, E and Sheldon, A "Duration and context effects on the perception of English /r/ and /l/: A comparison of Cantonese and Japanese speakers", *Language Learning, 36*(4), 505-21, 1986.
9. Jamieson, DG "Techniques for training difficult non-native speech contrasts", in K Elenius and P Branderud (Eds.) *Proceeds Internat Congress of Phonetic Sci, 4,* Stockholm: KTH and Stockholm University: Stockholm, 100-7, 1995.
10. Jamieson, DG and Morosan, DE "Training non-native speech contrasts in adults: Acquisition of the English /θ/-/ð/ contrast by francophones", *Percep & Psychophysics, 40*(4), 205-15, 1986.
11. Jamieson, DG, Ramji, K, Kheirallah, I and Nearey, TM "CSRE: The Canadian speech research environment", in JJ Ohala, TM Nearey, BL Derwing, MM Hodge and GE Wiebe (Eds.) *Proceeds Internat Conf on Spoken Lang Process*, (Edmonton: U. of Alberta), 1127-30, 1992.
12. Koo, JH "On the phonemic principles and romanization of Korean" in H. Sohn (Ed.) *The Korean Language: Its Structure and Social Projection*, 121-6, 1975.
13. Korean Overseas Information Service, *The handbook of Korea* (Seoul: Samhwa Printing), 8th ed., 1990.
14. Logan, JS, Lively, SE and Pisoni, DB "Training Japanese listeners to identify English /r/ and /l/: A first report", *J Acoust Soc of Amer, 89*(2), 884-86, 1991.
15. Mochizuki, M "The identification of /r/ and /l/ in natural and synthesized speech", *J Phonetics, 9,* 283-303, 1981.
16. Nábělek, AK and Donahue, AM "Perception of consonants in reverberation by native and non-native listeners", *J Acoust Soc of Amer , 72*(2), 632-4, 1984.
17. Pyun, K *Korean-Swedish Interlanguage Phonology* (Stockholm: Institute of Oriental Languages, U of Stockholm), 1987.
18. Sheldon, A and Strange, W "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception", *Appl. Psycholing., 3,* 243-61, 1982.
19. Strange, W "Phonetics of second-language acquisition: Past, present and future", in K Elenius and P Branderud (Eds.) *Proceeds Internat Congress of Phonetic Sci, 4,* Stockholm: KTH and Stockholm University: Stockholm, 76-83, 1995.
20. Takata, Y and Nábělek, AK "English consonant recognition in noise and in reverberation by Japanese and American listeners", *J Acoust Soc of Amer, 88*(2), 663-6, 1990.
21. Yamada, RA and Tohkura, Y "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners", *Percep & Psychophysics, 52*(4), 376-92, 1992.

Acknowledgements. This work was supported by the Natural Sciences and Engineering Research Council of Canada. We are grateful to Scott Lively, John Logan and David Pisoni for providing the stimuli used in our [r-l] test, and to Meg Cheesman for assistance implementing the ASRT and UWODFD tests. Correspondence should be addressed to jamieson@audio.hhcru.uwo.ca.