THE EFFECT OF TASK TYPE ON NATIVE SPEAKER JUDGMENTS OF L2 ACCENTED SPEECH

Ron I. Thomson and Paris Campagna

Dept. of Applied Linguistics, Brock University, St. Catharines, ON, Canada, L2S 3A1, rthomson@brocku.ca

1. INTRODUCTION

Research investigating adult second language (L2) accent often relies on human listeners to assess speech. While human judgments can measure intelligibility in absolute terms, they might be less precise for measuring fine-grained speaker-dependent variables. For example, there appears to be a relationship between lexical frequency [1] and familiarity [2], and L2 pronunciation accuracy. Some argue that these lexical effects do not reflect real differences in the speakers' articulation of more vs. less frequent words. Rather, judges may perceive the more frequent or familiar words to be more intelligible, because they are more frequent or familiar to the judges themselves [3]. We test this claim by measuring the effect of lexical familiarity on the intelligibility of L2 vowel productions judged in words, and as isolated vowels extracted from those words, thereby masking lexical context.

2. METHOD

2.1 Speakers

19 native speakers (NSs) of Canadian English (M age = 22.8, range = 19-49; 2 male, 17 female) provided baseline data. Beginner L2 English learners included 19 who spoke Standard Mandarin (M age = 40.1, range = 29-49; 4 male, 15 female) and 19 who spoke a Slavic L1 (M age = 38.6, range = 29-49; 12 female, 7 male), mostly Russian. The L2 speakers' mean length of residence in Canada was 15.6 months (range = 4 - 40).

2.2 Stimuli

Auditory elicitation stimuli were produced by a male speaker of Canadian English. They comprised 30 high frequency monosyllabic English verbs - three for each of ten Canadian English vowels: /i/, /□/, /e/, /□/, /æ/, /a/, /□/, /o/, /□/ and /u/ (e.g., 'keep', 'feed' and 'beat' for the vowel /i/).

2.3 Speaking task

In a quiet room, participants were given a written list of prompts, and heard each item played through headphones. They were recorded repeating each word immediately after hearing the auditory prompt, using a high quality digital recorder. Immediately after the speaking task, participants evaluated how familiar they were with each word they had just produced, using a 4-point scale where 0 = I don't know it; 1 = I might know it; 2 = I think I know it; and 3 = Yes, I know it.

2.4 Intelligibility Judgments

Using a sound-editing program, each recorded word was extracted from the original recording, normalized, and saved as a separate sound file. These files were then used to create a second set of files, in which the vocalic portion of each word was extracted, from the first glottal pulse after the initial consonant to the last glottal pulse before the following consonant. Perceptual screening was used to confirm that the selections were as accurate as possible.

Using *Praat* (www.praat.org), five phonetically trained judges assessed the intelligibility of each vowel production in three listening conditions. In Condition 1, all 5160 isolated vowel recordings were randomly presented to listeners, who were asked to identify the vowel they heard by clicking on one of ten buttons that represented the target categories. In Condition 2, the same procedure was used, except that listeners heard the entire word rather than the isolated vowel. The first 2 conditions were conducted in 26 equally sized blocks over several weeks. In Condition 3, recordings of the whole word were again presented over several weeks. However, items were blocked by word (e.g., all productions of 'keep' were presented before moving on to all productions of another word), and judges were asked to click on one of two buttons indicating that the target vowel was produced either correctly or incorrectly. For each condition, vowels were considered intelligible if 3 out of 5 judges agreed that it was a member of the intended category.

3. RESULTS

Words for each vowel category were assigned to one of three groups based on their mean familiarity scores, from 1, "most familiar" to 3, "least familiar". In one case, a tie in mean familiarity scores was broken through reference to its frequency in the British National Corpus. The NSs were perfectly familiar with all of the words. Hence, lexical familiarity rankings reflect only L2 learner ratings.

Among the 19 NSs who provided baseline production data, identification rates for isolated vowels varied from 60% to 100% (M=86%). In addition, 97% of the vowels extracted from the elicitation prompts were intelligible (one /u/ production was not). This confirms that it is possible for NS judges to identify vowels played in isolation.

Overall results by L1, listening condition and lexical familiarity ranks are shown in Figure 1.

Native speaker baseline intelligibility: A repeated measures ANOVA revealed significant main effects for listening

condition [$F(2,36) = 44.847$, $p < .000$, $\eta^2 = .714$] and lexical familiarity [$F(2,36) = 5.739$, $p = .007$, $\eta^2 = .242$], but not vowel. A significant interaction between listening condition and lexical familiarity was also found [$F(4,72) = 5.594$, $p = .001$, $\eta^2 = .237$]. Post-hoc Bonferroni adjusted t-tests indicated that mean intelligibility scores were significantly lower in the isolated vowel condition than in the other two listening conditions [$t(18) = -6.815$, $p < .000$; and $t(18) = -6.649$, $p < .000$ respectively]. Furthermore, mean intelligibility scores were significantly lower for the most familiar words, versus those that were least familiar [$t(18) = -3.172$, $p = .005$]. Three repeated measures ANOVAs, one for each listening condition, were conducted to investigate the interaction between listening condition and lexical familiarity. These revealed that the significant effect of lexical familiarity was only detectable in the isolated vowel condition [$F(2,36) = 6.187$, $p = .005$, $\eta^2 = .256$].
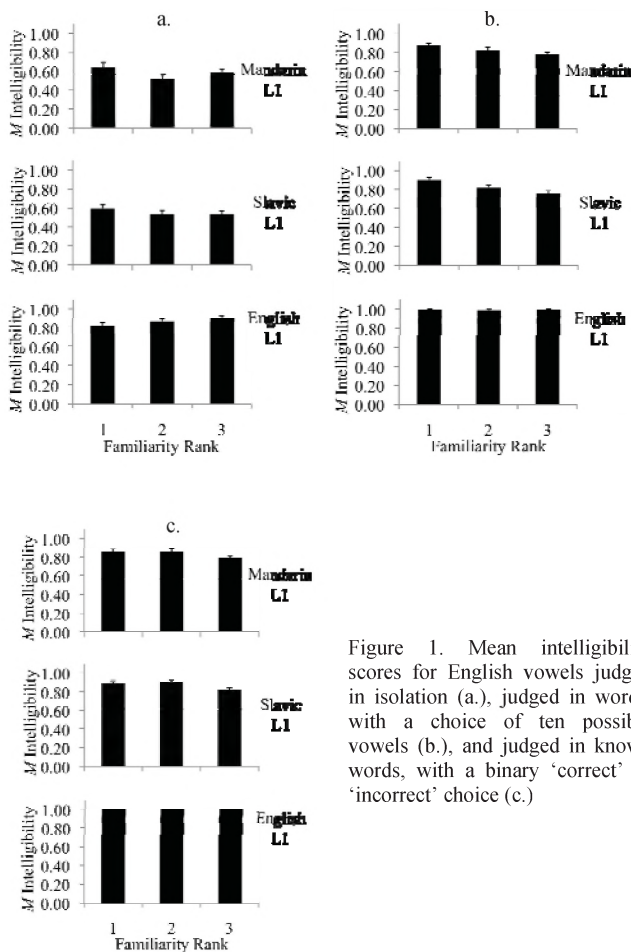

Figure 1. Mean intelligibility scores for English vowels judged in isolation (a.), judged in words, with a choice of ten possible vowels (b.), and judged in known words, with a binary 'correct' or 'incorrect' choice (c.)

L2 speaker intelligibility: A repeated measures ANOVA revealed significant main effects for listening condition [$F(2,72) = 169.997$, $p < .000$, $\eta^2 = .825$], lexical familiarity [$F(2,72) = 11.040$, $p < .000$, $\eta^2 = .235$], and vowel [$F(9,324) = 4.971$, $p < .000$, $\eta^2 = .121$]. In addition, significant interactions between listening condition and lexical familiarity [$F(4,144) = 5.124$, $p = .001$, $\eta^2 = .125$]; listening condition and vowel [$F(18,648) = 9.763$, $p < .000$, $\eta^2 = .213$]; and lexical familiarity and vowel [$F(18,648) = 5.059$, $p < .000$, $\eta^2 = .123$] were also found. There was no significant effect of L1. Post-hoc Bonferroni adjusted t-tests indicated that mean intelligibility scores were significantly lower in the isolated vowel condition than in the other listening conditions [$t(37) = -14.099$, $p < .000$; $t(37) = -13.172$, $p < .000$]. Condition 2 also resulted in lower scores than Condition 3 [$t(37) = -2.743$, $p = .009$]. In contrast to the results for NS productions, mean intelligibility scores were significantly higher for the most familiar words versus those that were second most and least familiar [$t(37) = 2.956$, $p = .005$; and $t(37) = 4.371$, $p < .000$ respectively]. There was no significant difference between vowel intelligibility in the second most and least familiar words. Three repeated measures ANOVA, one for each listening condition, revealed that the significant difference for lexical familiarity was detectable in all three listening conditions [$F(2,36) = 4.220$, $p = .018$, $\eta^2 = .105$; $F(2,36) = 15.80$, $p < .000$, $\eta^2 = .306$; and $F(2,36) = 9.691$, $p < .000$, $\eta^2 = .212$ respectively].

4. DISCUSSION

The results of this study indicate that lexical effects on the intelligibility of L2 vowel productions are present in the speech signal, and are not merely an artifact of listener bias towards evaluating vowels found in more familiar words as more intelligible. The intelligibility of NS vowel productions was also found to vary as a function of lexical context. However, while L2 vowels were most intelligible in the more familiar words, NS vowels were most intelligible in the less familiar words. This has implications for L2 speech learning, as it seems to indicate that the words that L2 learners are exposed to the most contain the poorest examples of the categories to be learned. Besides affecting speakers, lexical context also affects listeners. When they know what the intended category should be, listeners are more likely to evaluate the production as intelligible. This resulted in a ceiling effect for the NS productions, where real differences in the intelligibility of vowels, detected in the isolated vowel condition, were undetectable in listening conditions where lexical information was available.

REFERENCES
[1] Trofimovich, P., Gatbonton, E., & Segalowitz, N. (2007). A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition, 29*(3), 407-448.
[2] Thomson, R. I., & Isaacs, T. (2009). Within-category variation in L2 English vowel learning. *Canadian Acoustics, 37,* 138-139.
[3] Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America, 121,* 2327-2338.