SEGMENTAL ACCURACY AND L2 INTELLIGIBILITY: NATIVE AND NONNATIVE LISTENERS' PERSPECTIVES

XXX

XXX
XXX

ABSTRACT

This study investigated the contribution of listeners' L1 and stimulus properties to the perception of L2 speech. Thirty-nine listeners from five L1 backgrounds (native English, Korean, Mandarin, Indonesian, Turkish) completed a minimal-pairs forced choice task recorded by twenty Korean EFL talkers; the percentage of accurate word identification (i.e., intelligibility) was measured across listeners. Stimulus properties were measured by error coding focusing on ten target phonemes. Moderate to strong intergroup correlations suggested similarities in listeners' perceptions regardless of L1 backgrounds. No evidence for interlanguage speech intelligibility benefit (ISIB) was found for the nonnative listeners; in fact, native listeners always outperformed the nonnative listeners who shared (Korean) and did not share L1 with the talkers. Consonant errors (but not vowel errors) were linked to lower intelligibility. The effect of stimulus properties was further discussed in relation to the role of L1 Korean-specific interlanguage features in listeners' perception of L2 speech.

Keywords: intelligibility, pronunciation, segmental errors, listener factors, nonnative

1. INTRODUCTION

The importance of intelligibility over nativeness has become central to L2 pronunciation with the recognition that a speaker can be highly understandable despite a strong foreign accent [1]. To Munro and Derwing [2], intelligibility indicates listeners' actual understanding, whereas comprehensibility refers to listeners' perceived ease of understanding. Intelligibility has most commonly been measured using listeners' orthographic transcription, yet this study adopted a forced choice task, where listeners identify sounds to directly measure intelligibility in a way that precludes listeners from using information outside the speech signal, such as world and lexical knowledge [3].

An active line of intelligibility research has revealed that variability in intelligibility is linked to speakers' production accuracy in segmentals (e.g., [4]) and prosodic features such as stress, speech rate,

and rhythm (e.g., [5]). In addition to the speaker-related factors or *stimulus properties*, *listener-related factors* also play an important role in intelligibility, such as listeners' L1 [6]; expectations [7]; and familiarity with the speakers' accents [8]. Previous intelligibility research has generally focused on the perception of native listeners. Given the increasing L2 use among nonnative speakers globally, more research needs to consider nonnative listeners and different L1 backgrounds.

Some studies examined nonnative listeners' perceptions in terms of an interlanguage speech intelligibility benefit (ISIB), which refers to the benefit nonnative listeners receive when comprehending L2 speech better than native listeners. Several studies support this (e.g., [6, 9]). An ISIB was found for Mandarin listeners who were more accurate than native listeners at discriminating between a voiced or voiceless consonant (e.g., cub/cup) produced by Mandarin talkers on a forced-choice task [6]. Other studies found no support for an ISIB (e.g., [10, 11]). In [10], no evidence was found for Polish listeners listening to Polish-accented English speech on a transcription task. The Spanish listeners showing overall poorer performance indicated an interlanguage speech intelligibility *detriment* for nonnative listeners who differ in the L1 background from the nonnative talkers. Interestingly, no relationship was found between segmental or word stress errors and intelligibility, although segmental errors did affect the different perceptual measures studied.

Any investigation of listeners' understanding of L2 speech should consider both speaker-related (stimulus properties) and listener-related factors [12], but few studies have systematically examined stimulus properties in terms of segmental errors in relation to how they affect the intelligibility of L2 speech perceived by both native and nonnative listeners. Therefore, the current study aimed to address the following questions:

1. To what extent do native and nonnative English listeners share a response in their perception of L2 intelligibility?
2. What is the effect of the listener's L1 on the intelligibility of L2 speech?

3. What is the relationship between segmental errors and Korean learners' English intelligibility as perceived by native and nonnative English listeners?
    a. Which segmental sounds are most difficult for Korean learners to produce intelligibly as perceived by native and nonnative English listeners?

2. METHODOLOGY

2.1. Speech materials

The speech samples were the same as those used by [13] with native listeners. The speech samples were recorded by twenty native Korean speakers of L2 English in Seoul, South Korea; all speakers were in their final year at a high school at the time of the study (16 females, 4 males; Age = 15-17). Each speaker recorded their reading aloud ten target English words from a list of ten minimal pairs items (presented in Table 1). These minimal pairs were carefully chosen referring to a chapter on problematic phonemes specifically targeting Korean learners of English [14]. Recordings were embedded into each item of a survey on an online experiment platform *Gorilla* (https://gorilla.sc); items were randomized using the Randomise Trial function in Gorilla.

No.	Target		Foil	
1	b<u>i</u>t	/ɪ/	b<u>ea</u>t	/i/
2	<u>f</u>an	/f/	<u>p</u>an	/p/
3	la<u>gg</u>ing	/g/	la<u>ck</u>ing	/k/
4	<u>rh</u>yme	/r/	<u>l</u>ime	/l/
5	ri<u>pp</u>ing	/p/	ri<u>bb</u>ing	/b/
6	s<u>a</u>nd	/æ/	s<u>e</u>nd	/ɛ/
7	<u>sh</u>eet	/ʃ/	<u>s</u>eat	/s/
8	s<u>o</u>ng	/ɔ/	sung	/ʌ/
9	t<u>a</u>ste	/ey/	t<u>e</u>st	/ɛ/
10	<u>v</u>ote	/v/	<u>b</u>oat	/b/

Table 1: Minimal-pairs word list.

2.2. Listeners

Thirty-one nonnative English-speaking listeners were recruited at a large university in the Midwestern US from various academic disciplines. The participants were assigned to four listening groups based on their L1 background: Mandarin Chinese (n = 9), Indonesian (n = 7), Turkish (n = 9), and Korean (n = 6). All listener groups completed a language background questionnaire upon finishing the task. It was noted that both the Korean and the native English listener groups reported very high familiarity with Korean-accented English (self-rated; mean rating 4.67 and 4 out of 5, respectively).

2.3. Data collection procedure

Listeners completed a 200-item (ten recordings each from twenty Korean talkers) minimal-pairs forced-choice task administered using Gorilla. Listeners listened to each speech sample recorded by the speakers on each minimal-pairs forced-choice trial. The corresponding target-foil pair was then simultaneously displayed on the computer monitor, and listeners chose which word they heard. Listeners were allowed as long as they needed to respond. They pressed the play button when they were ready to hear the next speech sample. Listeners were given two short breaks throughout the task, which was completed in 25 minutes or less. Responses were scored automatically.

2.4. Data analysis

2.4.1. Listener task

The overall intelligibility scores were calculated by the mean percentage of words identified correctly by the listeners (henceforth, percentage correct) for each of the twenty Korean EFL speakers. For each of the ten target words, the mean percentage of items correctly identified was also computed across all twenty speakers (out of 620 items possible for each target word). The ten target words were then divided into those containing consonant targets (n = 6) and vowel targets (n = 4), and speakers' mean intelligibility scores for each category were also calculated. This led to a total of four types of intelligibility scores that were considered during analyses: overall intelligibility (i.e., mean percentage correct for each speaker), item intelligibility (i.e., mean percentage correct for each of the ten target items), consonant and vowel intelligibility (i.e., mean percentage correct for items containing consonant/ vowel targets).

2.4.2. Error coding

Each target phoneme in each sample (n = 200) was coded as either "on target" or "not on target", the latter being perceptually salient as indicating phoneme substitutions (e.g., /p/ for /f/ in 'faint') from General American (GA) English pronunciation. Note that phonetic errors (e.g., differences in aspiration, and vowel quality variations) that did not deviate substantially from GA English were disregarded and coded as "on target".

3. RESULTS

The results of the Cronbach alpha analyses confirmed the acceptable inter-rater reliability for the thirty-one

nonnative listeners' scores ($\alpha = 0.765$). All statistical analyses were carried out using software R and r effect sizes and eta-squared effect sizes were interpreted following [15] and [16], respectively.

3.1. Intergroup correlations

The intelligibility scores were averaged across all listeners separately for each of the nonnative listener groups, the distribution of which is presented in Figure 1. Table 2 summarizes the intergroup correlations on overall scores expressed as Pearson r values for all possible pairings of the five listener groups. The correlations were moderate to strong, ranging from .44 to .89.

Regarding consonant intelligibility, all the correlations between the groups were moderate to strong ($r = .58 - .85$). Vowel intelligibility revealed moderate to strong correlations for all ($r = .40 - .70$) but the Korean listener group, who showed nonsignificant and very weak correlations. In general, there were moderate to strong intergroup correlations on intelligibility, yet groups' correlations with the Korean listener group were somewhat weaker (and greatly weaker when it comes to vowel intelligibility) than with other groups.
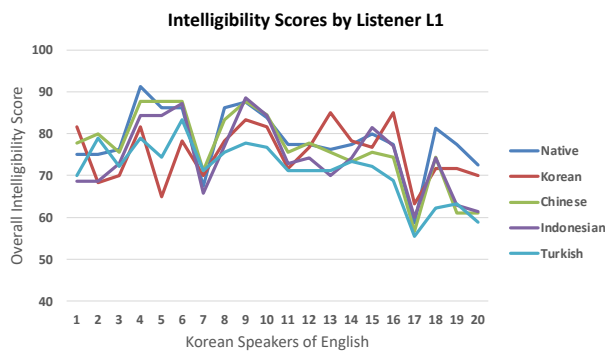

Figure 1: Mean intelligibility scores for the twenty Korean speakers of English by listener L1.

	EN	KR	CN	IN	TK
EN	1.00	–	–	–	–
KR	0.49 (.028)	1.00	–	–	–
CN	0.83 (<.001)	0.46 (.043)	1.00	–	–
IN	0.86 (<.001)	0.49 (.029)	0.88 (<.001)	1.00	–
TK	0.70 (.001)	0.44 (.050)	0.89 (<.001)	0.76 (.001)	1.00

Table 2: Intergroup correlations on the intelligibility scores.

3.2. Effects of listeners' L1 on intelligibility

To see the effect of listeners' L1 on intelligibility (see Figure 2), a one-way repeated measure ANOVA was performed using ez Package in R with listener L1 as a within-group factor (5 levels) on Korean speakers' intelligibility scores, and subsequently on consonant and vowel intelligibility. The results from the ANOVA analyses showed that there is extremely strong evidence of a difference between L1 listener groups in their intelligibility scores ($F(3, 48) = 7.482$, $p < .001$, $\eta^2_G = .09$) (Greenhouse-Geiser corrected). Post hoc (Bonferroni adjusted) analyses revealed that the Turkish listener group performed significantly lower than the native English and Chinese listener groups: native English ($M_{diff} = 7.23$, 95% CI [3.23, 11.22]), Chinese ($M_{diff} = 5.11$, 95% CI [2.13, 8.09]). The Indonesian listener group also performed significantly poorly on the task than the native listener group ($M_{diff} = 4.06$, 95% CI [0.91, 7.22]) ($p < .001$).

The effect of listener L1 was also statistical for consonant intelligibility ($F(4, 76) = 5.40$, $p < .001$, $\eta^2_G = .048$) and vowel intelligibility ($F(3, 49) = 4.47$, $p = .003$, $\eta^2_G = .11$). Regarding consonant intelligibility, post hoc analyses found that the native listener group performed significantly better than all the other groups: Chinese ($M_{diff} = 4.7$, 95% CI [0.3, 9.1]), Indonesian ($M_{diff} = 5.2$, 95% CI [1.0, 9.4]), Korean ($M_{diff} = 5.6$, CI [1.2, 10.0]), and Turkish ($M_{diff} = 6.6$, CI [2.2, 10.9]) ($p < .05$). No significant differences were found for any other L1 pairs (p-value approximates 1). Concerning vowel intelligibility, the Turkish listener group performed significantly poorer than the Chinese ($M_{diff} = 10.1$, $p < .001$, 95% CI [3.8, 16.4]) and the native English groups ($M_{diff} = 8.6$, $p = .017$, 95% CI [1.1, 16.0]).
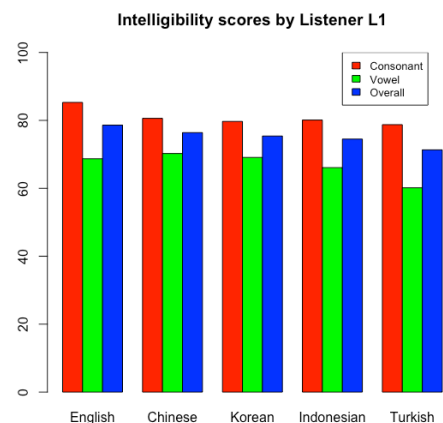

Figure 2: Intelligibility scores by five listener groups.

3.3. Segmental errors and intelligibility

When the overall intelligibility scores were compared with the number of target segment errors made by each speaker (M = 2.6, SD = 1.27, Range = 0-4; out of a maximum of 10 errors), moderate to strong negative associations were obtained for the native English (r = -.75), Chinese (r = -.56), and Indonesian (r = -.68) groups (p < .001). Slightly weaker but nonsignificant associations were found for the Korean (r = -.32, p = .164) and the Turkish group (r = -.28, p =.221). Consonant errors revealed moderate to strong negative correlations with overall intelligibility for all the listener groups ranging from -.79 to -.56 (p <.0025), whereas none of the correlations regarding vowel errors were significant (p > .05).

Table 3 shows item responses for native versus nonnative listeners combined on the listener task, along with the percentage error (identified in error coding) for different targets. Overall, native listeners and nonnative listeners showed similar responses to the items on the listener task. For both listeners, the three most intelligible target items were *sheet*, *lagging*, and *taste*, in which target segments /ʃ/, /g/, and /ey/ showed 0% error. The item *ripping* (/p/) also showed high intelligibility with 0% error. The three least intelligible items were *bit*, *sand,* and *vote*, where target segments /ɪ/, /æ/, and /v/ were of rather higher percentage error: 70%, 60%, and 25%, respectively. This is closely followed by the item *rhyme* (target consonant /r/ contrasting with /l/), which also showed lower item intelligibility with a 35% error rate.

Items	Listener-based intelligibility (%)		Production errors (%)
	Native	Nonnative	
b**i**t	36.88	53.23	70
fan	87.5	76.77	0
la**gg**ing	98.13	93.06	0
rhyme	68.13	70.48	35
ri**pp**ing	94.38	90.16	0
s**a**nd	50.63	54.84	60
sheet	98.75	90.16	0
s**o**ng	89.38	80.16	70
ta**s**te	97.5	93.55	0
v**o**te	64.38	48.39	25

Table 3: Item intelligibility by native vs. nonnative listeners; production errors on the right.

4. DISCUSSION

The questions posed at the beginning of this paper are now answered.

Listeners, regardless of their native status and L1, share a response to L2 speech in terms of intelligibility. On the other hand, the mean intelligibility scores for individual Korean EFL speakers varied greatly. It appears that the perception of L2 speech is more likely affected by the properties of the speech itself than the linguistic background of the listeners. Listener group comparisons showed no ISIB. In fact, native English listeners overall outperformed all the other listener groups. Rather, the overall poorer performance of the Turkish and the Indonesian listeners suggests an interlanguage speech intelligibility *detriment* for nonnative listeners who differ in the L1 background from the nonnative speakers.

Speakers with more segmental errors overall were found to be perceived as less intelligible by listeners; yet speakers' consonant accuracy, but not vowel accuracy, was negacorrelated with the speakers' overall intelligibility. Concerning the item intelligibility for individual target sounds, the finding of the present study showed that nonnative listeners found *sheet* (90.16%), *lagging* (93.06%), and *taste* (93.55%) the most intelligible, whereas *vote* (48.39%), *bit* (53.23%), and *sand* (54.84%) the least intelligible; this agrees with the native listener data [13]. Language-specific factors of the Korean language, such as having no distinction between lax and tense vowels (as in the case of b**i**t vs. b**ea**t, and s**e**nd vs. s**a**nd), and no equivalent sounds of labiodentals in the L1 (e.g., **v**ote) could account for Korean speakers' overall difficulty with the three items.

5. CONCLUSION

Intelligibility has predominantly been evaluated from the perspective of native English listeners responding to L2 speech, yet there is no valid reason to suggest native listeners' perceptions as basing the pedagogical solutions for teaching pronunciation. The results of this study found no essential differences between native and nonnative listeners in their perception of L2 speech. Concerning listener effects, no evidence for interlanguage speech intelligibility benefit was found for nonnative listeners. In this study, the nonnative listeners never outperformed the native English listeners in the forced-choice word identification task. In contrast, the properties of the L2 speech derived from the speakers' L1 phonological system appeared to be related to intelligibility. These findings should be refined by future research considering other listener and speaker factors related to intelligibility not examined in this study.

6. REFERENCES

[1] J. Levis, "Revisiting the Intelligibility and Nativeness Principles," *JSLP*, vol. 6, no. 3, pp. 310–328, Nov. 2020.

[2] M. J. Munro and T. M. Derwing, "Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners," *Language Learning*, vol. 45, no. 1, pp. 73–97, Mar. 1995.

[3] R. Thomson, "Measurement of accentedness, intelligibility, and comprehensibility," in *Assessment in Second Language Pronunciation,* O. Kang and A. Ginther, Eds., 1st ed. Routledge, 2017, pp. 11– 29.

[4] C. L. Rogers and J. Dalby, "Forced-Choice analysis of segmental production by Chinese-accented English speakers," *J Speech Lang Hear Res*, vol. 48, no. 2, pp. 306–322, Apr. 2005.

[5] O. Kang, R. I. Thomson, and M. Moran, "Which features of accent affect understanding? exploring the intelligibility threshold of diverse accent varieties," *Applied Linguistics*, vol. 41, no. 4, pp. 453–480, Aug. 2020.

[6] R. Hayes-Harb, B. L. Smith, T. Bent, and A. R. Bradlow, "The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts," *Journal of Phonetics*, vol. 36, no. 4, pp. 664–679, Oct. 2008.

[7] O. Kang and D. L. Rubin, "Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation," *Journal of Language and Social Psychology*, vol. 28, no. 4, pp. 441–456, Dec. 2009.

[8] B. Chung and H. K. M. Bong, "Intelligibility of Korean-accented English: Effects of listener's familiarity," *English Teaching*, vol. 76, no. 1, pp. 33–56.

[9] X. Xie and C. A. Fowler, "Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English," *Journal of Phonetics*, vol. 41, no. 5, pp. 369–378, Sep. 2013.

[10] I. A. Jułkowska and J. Cebrian, "Effects of listener factors and stimulus properties on the intelligibility, comprehensibility and accentedness of L2 speech," *JSLP*, vol. 1, no. 2, pp. 211–237, Oct. 2015.

[11] M. J. Munro, T. M. Derwing, and S. L. Morton, "The mutual intelligibility of L2 speech," *Stud. Sec. Lang. Acq.*, vol. 28, no. 01, Mar. 2006.

[12] B. Zielinski, "The Segmental/Suprasegmental Debate," in *The Handbook of English Pronunciation*, 1st ed., M. Reed and J. M. Levis, Eds. Wiley, 2015, pp. 397–412.

[13] I. Y. Na, "The impact of segmental accuracy on intelligibility," [Master's Scholarly Paper, University of Hawaiʻi at Mānoa]. ScholarSpace.

[14] P. Avery, S. Ehrlich, *Teaching American English pronunciation*. Oxford, England: Oxford University Press, 1992.

[15] L. Plonsky and F. L. Oswald, "How Big Is 'Big'? Interpreting effect sizes in L2 Research: Effect sizes in L2 Research," *Language Learning*, vol. 64, no. 4, pp. 878– 912, Dec. 2014.

[16] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Routledge, 2013.