

ASSESSING INTELLIGIBILITY OF L2 SPEECH BY NATIVE AND NON-NATIVE LISTENERS: A LONGITUDINAL STUDY

Alejandra Pesantez¹, Ronaldo Lima Jr²

¹University of Zurich, ²Federal University of Ceará
 alejandra.pesantezpesantez@uzh.ch; ronaldojr@letras.ufc.br

ABSTRACT

This study investigated the development of the English vowel contrast /i/-/ɪ/ by Ecuadorian learners over three semesters (T1, T2, T3). Development was examined by means of local intelligibility via native English, Swiss German, and Ecuadorian Spanish listeners' words transcribed correctly. It also explored the Interlanguage Speech Intelligibility Benefit (ISIB) between talkers and listeners from the same native language (L1) background. Speech samples collected from Ecuadorians (EC) with a Spanish L1 background (N=24) were presented to the three groups of listeners (N=15). Results of a Bayesian hierarchical model showed a tendency of improvement of the EC learners from T1 to T2, but not to T3. There was also evidence of ISIB for listeners, with EC listeners, in general, assessing a higher rate of intelligibility to EC speakers compared to native English and Swiss German listeners.

Keywords: Development, L2 speech, vowel, intelligibility, listeners.

1. INTRODUCTION

The majority of students in the world learn English as a second language (L2), in most cases in a formal instruction setting. In this kind of context, most L2 learners have limited access to native English speakers with more exposure to foreign-accented speech, receiving only a few hours of instructional input per week, with not much opportunity to use the target language outside of the classroom [1], [2]. Typical L2 learners of English in Ecuador, the population of interest in our study, are mainly exposed to Spanish-accented English during their years of classroom-based learning. This language environment can provide L2 speakers with different L2 phonological representations from those exposed to native English [3]. Furthermore, during the learning process of a language, it is common to observe some pronunciation errors, and segmental errors appear to be very common in L2 speech. Data from several studies suggest that some L2 pronunciation errors can be overcome during the first year of L2 immersion while others appear impervious [4], [5]. At the same time, results from previous

studies have proposed some hierarchies of difficulty in vowel production and perception based on the L1 of learners [4], [6]. For example, L1 Spanish speakers have difficulties producing tense/lax vowel differences [7]. Moreover, not all vowel contrasts are equally important for communication, and teachers should know which contrasts are problematic for L2 English speakers [8], [9]. For example, for L1 Spanish speakers, the English vowel /ɛ/ does not have a big impact on vowel intelligibility, even though it is sometimes pronounced as /e/ [4]. Conversely, the vowel contrast /i/ -/ɪ/ has a tremendous influence on intelligibility, and their mispronunciation can trigger misunderstandings. For this paper, intelligibility refers to "the extent to which a speaker's message is actually understood" [10]. Furthermore, the role of the listener in the intelligibility of speech is also crucial for effective communication [11], and the number of non-native L2 English listeners is more prevalent than that of native-English speakers (NEs).

Numerous studies on the assessment of L2 speech intelligibility have focused on native English (NE) listeners' ratings of vowels or word intelligibility [10]. However, there is some evidence to suggest that there is a benefit of speakers and listeners sharing a language, indicating that L2 speech with relatively low intelligibility for NE listeners could be highly intelligible for non-native English listeners who share the same L1 background [11], [12], [3]. For instance, the study by Uzun [12] demonstrated that the English words produced by Turkish speakers were more intelligible to Turkish listeners than NE listeners. This talker-listener relationship advantage was described by [13], [14] as the Interlanguage Speech Intelligibility Benefit (ISIB). In contrast to these findings, however, no evidence of ISIB was detected by other studies [15], [16]. For example, Munro et al. [15] did not find a consistent intelligibility benefit for languages shared between listeners with Cantonese, Japanese, Mandarin, and English backgrounds and speakers of Cantonese, Japanese, Polish, and Spanish. The explanation for these conflicting results might be due to the different kinds of elicitation tasks for measuring L2 intelligibility (global vs local) used, as well as the non-native listeners' proficiency levels.

The goal of the present study is two-fold. Firstly, we aimed to track the early stages of L2 development

of the English vowel contrast /i/-/ɪ/ produced by 24 Ecuadorian speakers. Development was analyzed by means of local intelligibility to investigate L2 learning processes and speaker errors that lead to problems for listeners [17]. Secondly, we aimed to investigate ISIB via native English, Swiss German, and Ecuadorian listeners' transcriptions of EC and NE speakers' recorded words.

2. METHOD

2.1. Speakers

This data is part of a longitudinal study about foreign language acquisition conducted by the first author and used in different publications with different purposes. The study started with 24 Spanish-speaking Ecuadorian learners of English (EC) aged between 18 and 28, but for the following two recordings only 21 participants showed up. The EC speakers had not lived in an English-speaking country for the purpose of studying and are L1 Spanish monolinguals. However, three participants reported having studied French, Portuguese, and Quechua respectively, reaching a basic level in these languages. Participants were majoring in English language teaching at a state university in Ecuador. They were in the third, fourth, and fifth levels of their studies in the three recordings (T1, T2, and T3) that were analyzed. In the fourth semester, they studied phonetics, and in the fifth phonology. They also had experience learning English for four hours per week during their secondary school in Ecuador with non-native English input most of the time. Furthermore, in the English program at the university, most of the English instructors were non-native English speakers. All participants gave their written informed consent for each recording session and were paid for their participation. Furthermore, 4 American speakers of English (NE) (mean age = 19.75), were used as a control group for comparative purposes in the intelligibility assessment. Participants in the NE group were two female and two male students from different parts of the United States of America, who were in Ecuador studying Spanish as a second language. They were recorded at T1, and their participation was voluntary.

2.2. Speech production

The EC and NE speakers produced 40 isolated monosyllabic words containing the following English vowel contrasts /i:/-/ɪ/; /u:/- /ʊ/; /ɛ /-/æ/; /ʌ/-/ ʌ/) in a CVC or CVCC context through a picture-naming task. For this study, we only present words containing the /i/-/ɪ/ contrast because it has a high functional load in English, such as the word pairs of this research:

cheap-chip, keys-kiss, feet-fit, seat-sit, and sheep-ship [18]. Second, previous studies have suggested that speakers from diverse L1 backgrounds (e.g., Catalan, Cantonese) have difficulties to produce the /i/-/ɪ/ pairs. [6], [7], [19]. Participants were recorded at six months intervals, upon finishing each semester.

The recordings were conducted in the radio station of a state university in Ecuador, and the audio was captured using a Zoom H2n handy recorder at a 44.1 kHz sampling rate, and a 16-bit quantization rate. Before starting the recordings, participants were familiarized with the task in a preliminary trial, with instructions written in English and appearing on the first slide. No recording was performed during the preliminary phase. All the pictures were presented in a random order with their Spanish translation next to them to avoid the effect of orthography [4] in the production of the English segments. Participants repeated each word twice to avoid mispronunciation or hesitation during the first production. They were asked to say the words in a natural way and at a normal volume. NE speakers followed the same procedures.

2.3. Listeners

Fifteen listeners participated in the current study. Most of them live in Ecuador or were visiting the country at the moment of the judgement sessions. 5 were American speakers of English (NEL) aged between 22 and 45, two of whom had lived in Ecuador for over 4 years, and two lived in Central America for the purpose of studying Spanish for a mean of 1.65 years. There were also 6 Ecuadorian (two English teachers, one tourist guide and three university students) and 4 Swiss German listeners, who will be referred to as ECL and GL respectively, and who were fluent in English. Most of the GL were English teachers (primary, secondary, or university) in their respective countries. The members of the ECL group were aged between 22 and 38. Two of the participants had studied in an English-speaking country (mean=1 year). The members of the GL group were aged between 26 and 45. Three of them were Swiss German speakers and one was a German speaker fluent in Swiss German. All the GL reported having also studied French. One GL had lived in Ecuador for around 9 years, being fluent in Spanish as well. Prior to participating, they declared not to have hearing problems. A language questionnaire was administered to each listener prior to the listening tasks. Their participation was voluntary.

2.4. Procedure

Participants were seated in front of a computer monitor using headphones in a quiet room during the

sessions. Using Praat software [20], each listener pressed a button to play the target word and heard the randomly ordered presentations of the stimuli twice through headphones. They heard the words containing the /i/-/ɪ/ vowel contrast produced by the EC during the three semesters and the words produced by NE at T1, T2, and T3 which were presented during the three rating sessions as a control group. They were instructed to write out exactly what was said using a computer keyboard. Before starting the task, listeners were trained with four trials and regulated the loudness of the stimuli. Each session lasted from 45 to 50 minutes.

2.5. Statistical analyses

EC and NE individual words were scored as correct or incorrect based on whether orthographic transcriptions by listeners matched the target words produced by the speaker [21]. This data was used to fit a Bayesian mixed-effects logistic regression model, with *vowel* ([i] or [ɪ]), *time* (T1, T2 and T3), *speaker nationality* (EC or NE) and *listener nationality* (ECL, NEL and GL) as predictor variables, with varying intercepts for word and listener, and varying slopes for speaker as a function of time.

3. RESULTS

3.1. Aim 1: To track early stages of L2 development of English vowel contrast /i/-/ɪ/ produced by 24 Ecuadorian speakers.

We fitted a Bayesian logistic regression model¹ (with flat priors) to the data using the brms package [22] for R, whose coefficients are presented in Table 1. Instead of providing point estimates with the probability of the data given the null hypothesis (i.e., p-values), Bayesian models provide probability distributions of the parameters given the data. Figure 1 presents the graphs for the probability distributions from Table 1. The corresponding measurement of error model yields an estimated posterior mean of 0.37, with a 95% credible interval [0.12, 0.65] showing the increase in the probability of correct vowels for T2 in comparison with T1 (the intercept), but not so much for T3 in comparison with T2 with an estimated posterior mean of 0.48, with 95% credible interval [0.23, 0.73]. It also shows that the model predicts a credible higher probability of correct vowels for NE speakers.

Predictors	correct.code	
	Log-Odd CI (5%)	
Intercept	0.11	-0.50 – 0.71
speaker_nationality:NE	0.72	0.10 – 1.37
time: T2	0.37	0.12 – 0.65
time: T3	0.48	0.23 – 0.73
vowel: ɪ	-0.10	-0.81 – 0.67
listener_nationalityGL	-0.22	-0.62 – 0.20
listener_nationalityNEL	-0.22	-0.60 – 0.17
listener_nationalityNE:listener_GL	1.33	0.76 – 1.95
listener_nationalityNE:listener_NEL	2.43	1.59 – 3.54
timeT2: listener_nationalityGL	0.16	-0.09 – 0.42
timeT3: listener_nationalityGL	0.41	0.15 – 0.69
timeT2: listener_nationalityNEL	0.24	-0.03 – 0.53
timeT3: listener_nationalityNEL	0.46	0.17 – 0.76
vowel: listener_nationalityGL	0.64	0.43 – 0.85
vowel: listener_nationalityNEL	0.61	0.40 – 0.83

Table 1: Coefficients for Regression Model.

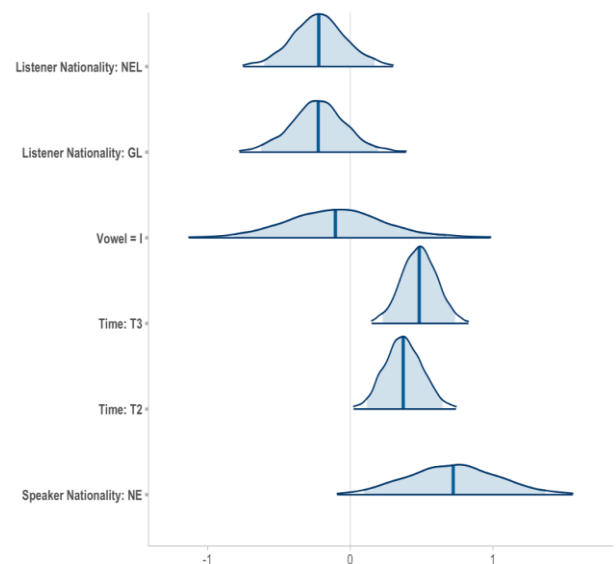


Figure 1: Posterior probability distributions of the model's coefficients.

3.2. Aim 2: To investigate ISIB via native English, Swiss German, and Ecuadorian Spanish listeners' transcriptions of Ecuadorian learners' words.

The effects of vowel and listener nationality, however, were not found to be credible. For vowel, the corresponding measurement error model yields an estimated posterior mean of -0.10, with 95% credible interval [-0.81, 0.67] (with parts of the distribution in the negative and positive sides). For GL, the corresponding measurement error model yields an

¹ Model: correct.vowel ~ speaker_nationality + time + vowel + listener_nationality + speaker_nationality:listener_nationality + time:listener_nationality + vowel:listener_nationality + (time|ID) + (1|word) + (1|listener)

estimated posterior mean of -0.22, with 95% credible interval [-0.62, 0.20]. Similarly, for NEL the corresponding measurement error model yields an estimated posterior mean of -0.22, with 95% credible interval [-0.60, 0.17] though the latter yielded some insightful interactions for speaker nationality and time (see Figs 3a, 3b,) respectively. Figure 3a shows that ECL listeners rated EC speakers slightly higher than the other listeners, but rated NE speakers much lower than the other listeners. Finally, Fig 3b shows that ECL listeners assigned higher rates for /i/ and lower rates for /ɪ/ when compared to the other listeners.

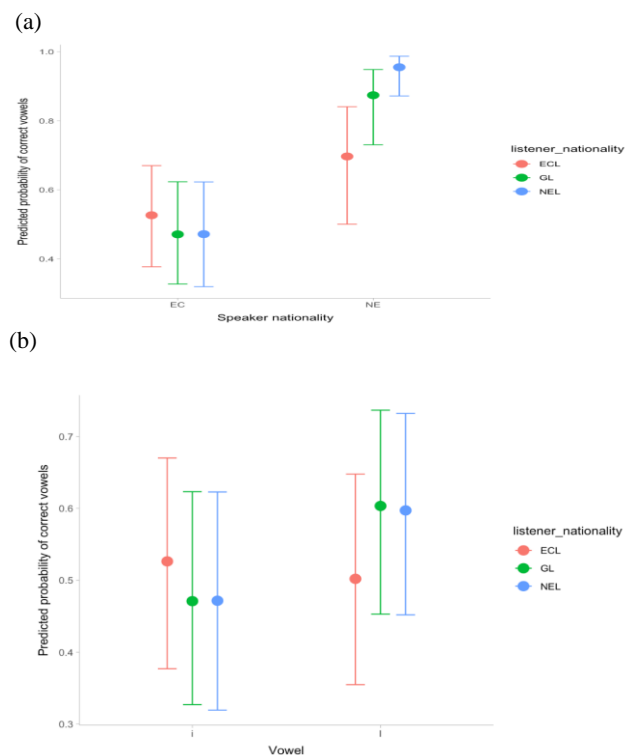


Figure 2: Probabilities predicted by the model.

4. DISCUSSION

Several reports have shown that L2 learners can accurately produce some L2 phonemes over time [4]. In our study, EC speakers started with clear lower proportions of correct vowels and tended to improve as time increased. The biggest improvement was observed from T1 to T2 and with increased probability of correct vowels. One possible explanation for this pronunciation improvement could be, apart from exposure, the positive effect of the phonetics classes, which might lead L2 speakers to establish representations of the vowel categories in long-term memory. From T2 to T3 little increased probability of correct vowel was observed. It could be possible that L2 speakers showed an initial rapid

improvement of the phonetic learning which followed a plateau [9]. Previous studies have shown that L1 Spanish speakers have difficulties differentiating tense-lax vowel contrasts. However, in our study, EC speakers seemed to have slightly better proportions for lax [ɪ] than for tense [i].

The second aim was to investigate ISIB via native English, Swiss German, and Ecuadorian Spanish listeners' transcriptions of Ecuadorian and NE speakers' words. In general, a benefit between listeners and speakers who share the same L1 background was slightly observed, with non-native speakers being more intelligible to ECL than to GL and NEL, though with no robust credibility. This finding goes in the direction of Uzun's [12], in which a benefit from an L1 shared between Turkish listeners and speakers was observed on intelligibility through a transcription task. We also found that NE speakers got the lowest intelligibility score by the ECL. This result could be related to the ECL experience since three of the participants did not use English every day compared to the others. Moreover, ECL, GL and NEL differed on how they perceived word intelligibility produced by the EC speakers. ECL could hear more accurately than NE listeners words containing the vowel /i/. However, GL and NEL were more accurate when transcribing words containing /ɪ/. A possible explanation could be the differences between the acoustic information that NEL and GL rely on when listening to NE speech and the information available in the Spanish-accented English speech.

5. CONCLUSIONS

The small listeners' sample size did not allow to conclude an ISIB on talker-listener relationship advantage. However, we consider that our study contributes data to an important line of understandings listeners' perception of word intelligibility on the /i/-/ɪ/ vowel contrast and EC speakers' improvement over time. For future research, it is important to see different listening conditions and different methods of evaluation including a greater number of native and non-native listeners. Moreover, it is difficult to conclude that ECL cannot distinguish well the NE speakers' vowels due to the lower number of NE speakers than the EC speakers. However, this study should be replicated with homogeneous number of native and non-native speakers to observe if a poor vowel intelligibility rating by non-native listeners is the result of the lack of distinguishing tense/lax contrast. Finally, the study should be repeated with L2 speakers' English teachers rating their students' utterances to provide more insights of the familiarity of the L1 speaker's background.

6. REFERENCES

- [1] C. Muñoz, “Contrasting effects of starting age and input on the oral performance of foreign language learners,” *Appl Linguist*, vol. 35, no. 4, pp. 463–482, Sep. 2014, doi: 10.1093/applin/amu024.
- [2] C. Muoz, “Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning,” *Appl Linguist*, vol. 29, no. 4, pp. 578–596, Dec. 2008, doi: 10.1093/applin/amm056.
- [3] X. Xie and C. A. Fowler, “Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English,” *J Phon*, vol. 41, no. 5, pp. 369–378, 2013, doi: 10.1016/j.wocn.2013.06.003.
- [4] J. E. Flege, O.-S. Bohn, and S. Jang, “Effects of experience on non-native speakers’ production and perception of English vowels,” *J Phon*, vol. 25, no. 4, pp. 437–470, Oct. 1997, doi: 10.1006/JPHO.1997.0052.
- [5] W. Baker and P. Trofimovich, “Perceptual paths to accurate production of L2 vowels: The role of individual differences,” *IRAL - International Review of Applied Linguistics in Language Teaching*, vol. 44, no. 3, pp. 231–250, Nov. 2006, doi: 10.1515/IRAL.2006.010.
- [6] J. C. Mora and N. Fullana, “Production and perception of English /i:/-/ɪ/ and /ae/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age,” 2007. [Online]. Available: www.icphs2007.de
- [7] J. Cebrian, “OLD SOUNDS IN NEW CONTRASTS: L2 PRODUCTION OF THE ENGLISH TENSE-LAX VOWEL DISTINCTION.” [Online]. Available: www.icphs2007.de
- [8] J. M. Levis, “What Should and Should Not Be Taught,” in *Intelligibility, Oral Communication, and the Teaching of Pronunciation*, Cambridge University Press, 2018, pp. 239–264. doi: 10.1017/9781108241564.013.
- [9] J. M. Levis, “Changing Contexts and Shifting Paradigms in Pronunciation Teaching,” 2005. [Online]. Available: <https://www.jstor.org/stable/3588485?seq=1&cid=pdf->
- [10] J. E. Flege, M. J. Munro, and I. R. A. MacKay, “Factors affecting strength of perceived foreign accent in a second language,” *J Acoust Soc Am*, vol. 97, no. 5, pp. 3125–3134, 1995, doi: 10.1121/1.413041.
- [11] P. Lochland, “Intelligibility of L2 speech in ELF,” *Australian Journal of Applied Linguistics*, vol. 3, no. 3, pp. 196–212, Dec. 2020, doi: 10.29140/ajal.v3n3.281.
- [12] T. Uzun, “The Interlanguage Speech Intelligibility Benefit for Turkish Speakers of English,” *Dil Eğitimi ve Araştırmaları Dergisi*, Apr. 2022, doi: 10.31464/jlere.1024616.
- [13] T. Bent and A. R. Bradlow, “The interlanguage speech intelligibility benefit,” *J Acoust Soc Am*, vol. 114, no. 3, pp. 1600–1610, Sep. 2003, doi: 10.1121/1.1603234.
- [14] 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,” *J Phon*, vol. 36, no. 4, pp. 664–679, Oct. 2008, doi: 10.1016/j.wocn.2008.04.002.
- [15] M. J. Munro, T. M. Derwing, and S. L. Morton, “THE MUTUAL INTELLIGIBILITY OF L2 SPEECH,” *Stud Second Lang Acquis*, vol. 28, no. 01, pp. 111–131, Mar. 2006, doi: 10.1017/S0272263106060049.
- [16] R. C. Major, S. F. Fitzmaurice, F. Bunta, and C. Balasubramanian, “The Effects of Nonnative Accents on Listening Comprehension: Implications for ESL Assessment,” 2002.
- [17] M. J. Munro and T. M. Derwing, “Intelligibility in research and practice: Teaching priorities,” in *The Handbook of English Pronunciation*, Wiley, 2015, pp. 375–396. doi: 10.1002/9781118346952.ch21.
- [18] A. Sewell (2017). “Functional load revisited: Reinterpreting the findings of ‘lingua franca’ intelligibility studies”. *Journal of Second Language Pronunciation*, vol. 3, no 1, pp. 57-79, 2017, doi: 10.1075/jslp.3.1.03sew
- [19] M. J. Munro, “On the difficulty of defining “Difficult” in second-language vowel acquisition”. *Frontiers in Communication*, 6. Aug. 2021, doi: 10.3389/fcomm.2021.639398
- [20] Boersma, P. Weenink, D. (1992). *PRAAT: Doing phonetics by computer*. (6.1.02) [Computer software]. <http://www.praat.org/>
- [21] M. J. Munro and T. M. Derwing, “Foreign accent, comprehensibility, and intelligibility in the speech of second language learners,” *Lang Learn*, vol. 49, no. PART 1, pp. 285–310, 1999, doi: 10.1111/0023-8333.49.s1.8.
- [22] H. Badr. P.Bürkner Paul, (2021) “*bayesian: Bindings for Bayesian TidyModels.*,” doi:10.5281/zenodo.6654386, <https://hsbadr.github.io/bayesian/>.