

Asynchronous online peer judgments of intelligibility: simple task, complex factors

Suzanne M. Yonesaka¹

Abstract. Pronunciation learners can benefit from peer feedback in a Computer-mediated Communication (CMC) environment that allows them to notice segmentals and suprasegmentals. This paper explores the intelligibility judgments of same-L1 peers using P-Check (Version2, <https://ver2.jp>), a Learning Management System (LMS) plug-in that aggregates peer feedback on local intelligibility (Munro & Derwing, 2015). P-Check randomly delivers written prompts for learners to record. Recordings are randomly delivered to peers who choose from a drop-down menu which utterance was perceived. Aggregated judgments from peers and from the instructor are displayed to learners as feedback on intelligibility. This study used eight segmental contrasts: /b-v/, /s-θ/, /l-l/, /l-ɹ/-clusters, /æ-ʌ/, /ɑ-ʌ/, /ɑ-ov/, and /i-i/. Participants (N=38) made 3,451 intelligibility judgments on 1,203 recordings. The effects of rater listening discrimination proficiency and of utterance intelligibility were examined in six contrasts using Generalized Estimating Equations (GEE). Results showed that intelligibility was generally a significant predictor of judgment accuracy, but rater listening discrimination proficiency was not.

Keywords: pronunciation, intelligibility, peer feedback.

1. Background

In pronunciation learning, the effectiveness of same-L1 peer feedback in English as a foreign language environments has yet to be fully explored. There is some evidence that same-L1 learners can benefit from peer pronunciation feedback, especially in asynchronous CMC environments that allow repeated listening to recorded speech, giving learners time to notice pronunciation features (Correa & Grim, 2014; Gilakjani, Ahmadi, & Ahmadi, 2011). However, feedback

1. Hokkai-Gakuen University, Sapporo, Japan; suzanne@hgu.jp; <https://orcid.org/0000-0002-2534-1632>

How to cite this article: Yonesaka, S. M. (2019). Asynchronous online peer judgments of intelligibility: simple task, complex factors. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thoušný (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 407-412). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.1045>

from same-L1 learners may be problematic in task-based communication due to learners converging on a shared non-standard pronunciation. Walker (2005) recommends preventing convergence through highly-controlled activities. Thus, the present study directs participants' feedback to selected features by having them make a forced-choice judgment of local intelligibility (Munro & Derwing, 2015).

Pronunciation instruction would benefit from a better understanding of what factors underlie the accuracy of these same-L1 peer judgments of intelligibility. This exploratory study focuses on two aspects: the stimulus and the learner. The research questions are: (1) to what extent does the accuracy of local intelligibility judgments by same-L1 peers vary depending on the targeted phoneme and utterance accuracy, and (2) to what extent does it depend on rater listening discrimination ability?

2. Method

2.1. Participants

The 38 participants ($M=17$, $F=21$) in this convenience sample were Japanese university students enrolled in an elective first-year practical English phonetics course who provided their informed consent following Teaching English as a Second Language (TESOL) standards.

2.2. Classroom environment

The language targets were eight segmental contrasts that are difficult for this learner population: /b-v/, /s-θ/, /l-l/, /l-l/-clusters, /æ-ʌ/, /ɑ-ʌ/, /ɑ-oʊ/, and /i-i/. Materials consisted of 47 pairs of two-line contrastive conversations with L1 glosses. The first line of each conversation differed in one phoneme, such as:

Conversation 1. A: *He is a good leader.* B: *Everyone trusts him.*

Conversation 2. A: *He is a good reader.* B: *He loves books.*

After receiving focused instruction on the targeted phoneme, participants did individual online listening discrimination and pronunciation practice. Due to time constraints, this practice was not completed for the /s-θ/ or /ɑ-oʊ/ contrast.

The learning sequence included a listening discrimination pre-test, shadowing, listening discrimination practice, visual input, pronunciation practice, and choral repetition of the contrastive conversations to familiarize participants with their meaning and pronunciation. Finally, participants engaged in online peer judgments of intelligibility for approximately 15 minutes per contrast.

2.3. Software

The peer judgments were conducted using P-Check, a plug-in for Glexa, a proprietary LMS that has been used by more than 100,000 students in over 1,000 university courses throughout Japan. P-Check randomly presented the first line of one of the two conversations onscreen for the learner to record. Recordings were randomly delivered to peers who selected the appropriate second line of the conversation from a drop-down menu. After recordings received four judgments, they were taken out of circulation. A Native-Speaker (NS) rater also used P-Check to judge the intelligibility of all recordings. Peer and NS rater feedback for each recording was displayed to individual participants.

2.4. Data collection and analysis

Data were gathered from the P-Check database. Data consisted of participants' intelligibility judgments ($n=3,451$) which were compared to the NS rater's judgments of the 1,215 recordings produced by participants.

In further analysis, the relative effects of utterance intelligibility and rater listening discrimination proficiency were modeled for the six contrasts listed in [Table 3](#). GEE, which produces a population-average model, was used because it can accommodate repeated categorical outcomes while accounting for a different number of outcomes per participant.

The model included one centered covariate, listening discrimination proficiency, and one factor, intelligibility of the utterance. The events-in-trials outcome variable was accurate/inaccurate judgment of intelligibility. The models used binomial distribution with a logit-link function, exchangeable structures, and robust standard errors ([Heck, Thomas, & Tabata, 2012](#)). Models were run separately for each contrast and those with the lowest QICC², a criterion used in model selection, were chosen. An alpha level of .05 was used for all statistical tests.

2. Quasi likelihood under independence model criterion

3. Initial results and discussion

Research Question 1 asks to what extent the accuracy of local intelligibility judgments vary depending on the targeted phoneme and utterance accuracy. Table 1 suggests that participants were more likely to make accurate judgments when the utterance was intelligible (to the NS rater) than when it was not.

Table 1. Accuracy of peer judgments by perceived intelligibility

Peer judgment	Accurate (%)	Inaccurate (%)	Total (%)
Intelligible	1,819 (52.7)	437 (12.7)	2,256 (65.4)
Unintelligible	460 (13.3)	735 (21.3)	1,195 (34.6)
Total	2,279 (66.0)	1,172 (34.0)	3,451 (100)

Of all judgments, 21.3% occurred when peers were unable to recognize an intelligible phoneme; these may be related to low listening discrimination ability. Only 12.7% of judgments involved participants judging an unintelligible phoneme to be intelligible; these judgments may involve using knowledge of the L1 phonology.

Table 2 indicates substantial variation in mean intelligibility among the contrasts. The least intelligible contrasts were /s-θ/ and /ɑ-ɔʊ/, which had been taught but not fully practiced, and /l-ɹ/ clusters. The /æ-ʌ/ and /ɑ-ʌ/ contrasts were most intelligible and were judged most accurately.

Table 2. Accuracy of peer intelligibility judgments by contrast

Contrast	Intelligibility (NS)		Peer Judgments	Peer Accuracy	
	M	SD	n	M	SD
/b-v/	.80	.401	334	.65	.479
/s-θ/	.72	.452	426	.66	.473
/l-ɹ/	.73	.445	342	.64	.480
/l-ɹ/ clusters	.68	.466	395	.59	.493
/i-ɹ/	.73	.444	480	.71	.453
/ɑ-ɔʊ/	.65	.476	627	.63	.484
/ɑ-ʌ/	.80	.403	429	.70	.457
/æ-ʌ/	.86	.346	418	.69	.463
Total			3,451		

Research Question 2 asks to what extent the accuracy of judgments depends on rater listening discrimination ability. This was measured for six contrasts by the listening discrimination pre-test at the beginning of the teaching sequence. Participants

heard 20 pairs of words, half minimal pairs (e.g. *lake*, *rake*) and half tokens of the same word (e.g. *lake*₁, *lake*₂), and they marked each pair same or different. Table 3 indicates that participants discriminated /æ-ʌ/ the best and /l-ɪ/-clusters the least well.

Table 3. Listening discrimination scores by contrast

	n	Min.	Max.	M	SD
/b-v/	30	11	18	14.33	1.77
/l-ɪ/	37	10	19	14.05	2.25
/l-ɪ/ clusters	32	6	17	11.44	2.41
/i-ɪ/	31	15	20	18.29	1.53
/ɑ-ʌ/	32	14	20	17.25	1.83
/æ-ʌ/	31	15	20	18.97	1.25

Results of the GEE analysis were as follows. For the /b-v/ and /l-ɪ/-clusters contrast, neither intelligibility nor listening discrimination were significant predictors of judgment accuracy. For the remaining contrasts, parameter estimates showed that unintelligible utterances received accurate intelligibility judgments at a significantly lower rate than intelligible utterances (/l-ɪ/: Wald $\chi^2(1)=6.054$, $p=.014$, $\beta=-.584$); /i-ɪ/: Wald $\chi^2(1)=12.388$, $p<.001$, $\beta=-.949$; /æ-ʌ/: Wald $\chi^2(1)=11.158$, $p=.001$, $\beta=-.928$; /ɑ-ʌ/: Wald $\chi^2(1)=69.707$, $p<.001$, $\beta=-2.979$). For /ɑ-ʌ/, listening discrimination was also a predictor of judgment accuracy (Wald $\chi^2(1)=5.888$, $p=.015$, $\beta=.141$).

4. Conclusion

Intelligibility was a significant predictor of judgment accuracy, except for /b-v/ and /l-ɪ/-clusters. Closer examination reveals further variation even within some contrasts. For example, /b/ had 75% judgment accuracy while /v/, a phoneme not in the participants' L1 inventory, had only 56% accuracy. A different pattern was seen for /l/ and /ɪ/, both of which were judged with 61% mean accuracy. However, the /l-ɪ/ contrast showed strong variation at the item level, with accuracy ranging from 25% (*long*) to 91% (*lamp*).

Although a detailed analysis is beyond the scope of this paper, it is clear that intelligibility judgments were highly sensitive to target variability. Unexpectedly, listening discrimination ability was found to predict judgment accuracy only for one contrast, possibly indicating that a more robust measure of this covariate is needed.

References

- Correa, M., & Grim, F. (2014). Audio recordings as a self-awareness tool for improving second language pronunciation in the phonetics and phonology classroom: sample activities. *Currents in Teaching & Learning*, 6(2), 55-63. <https://www.worcester.edu/Currents-Archives/>
- Gilakjani, A., Ahmadi, S., & Ahmadi, M. (2011). Why is pronunciation so difficult to learn? *English Language Teaching*, 4(3), 74-83. <https://doi.org/10.5539/elt.v4n3p74>
- Heck, R. H., Thomas, S., & Tabata, L. (2012). *Multilevel modeling of categorical outcomes using IBM SPSS*. Routledge. <https://doi.org/10.4324/9780203808986>
- Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed and J. M. Levis (Eds.), *The handbook of English pronunciation*, (pp. 377–396). Malden, MA: Wiley Blackwell. <https://doi.org/10.1002/9781118346952.ch21>
- Walker, R. (2005). Using student-produced recordings with monolingual groups to provide effective, individualized pronunciation practice. *TESOL Quarterly*, 39(3), 550-558. <https://doi.org/10.2307/3588495>



Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

© 2019 by Editors (collective work)
© 2019 by Authors (individual work)

CALL and complexity – short papers from EUROCALL 2019
Edited by Fanny Meunier, Julie Van de Vyver, Linda Bradley, and Sylvie Thouéšny

Publication date: 2019/12/09

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2019.38.9782490057542>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Cover theme by © 2019 Frédéric Verolleman
Cover layout by © 2019 Raphaël Savina (raphael@savina.net)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-54-2 (Ebook, PDF, colour)

ISBN13: 978-2-490057-55-9 (Ebook, EPUB, colour)

ISBN13: 978-2-490057-53-5 (Paperback - Print on demand, black and white)

Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2019.