Automated Feedback and Individualised Training System for L2 Speech

Abstract

While scholars emphasise the importance of intelligibility over nativelikeness in second language (L2) education, the mechanisms underlying understanding between speakers and listeners remain unexplored. Despite advancements in automated speech recognition (ASR) that hold great potential for L2 teaching, few have examined how ASR can be optimised for different types of classroom learners. Utilising machine learning techniques, we aim to develop an assessment tool to model how listeners from diverse backgrounds intuitively evaluate the intelligibility of L2 speech. A subsequent intervention study will investigate the efficacy of the feedback provided by the tool for learners with varied aptitude, motivation, and beliefs.

Investigators

Kazuya Saito (PI)
Professor, Institute of Education, University College London

Nobuaki Minematsu (Co-I)
Professor, School of Engineering, University of Tokyo

Grant Duration
September 2025 (3 years)

Grant Amount
£310,535

Automated Feedback and Individualised Training System for L2 Speech

Project

In today's globalised world, one's degree of L2 speaking proficiency has direct implications for university admissions, employability, and cultural understanding. Experts in education frequently emphasise the role of intelligibility over nativelikeness.[1] The bidirectional model of comprehension further posits that successful understanding is influenced by both the speaker's linguistic capabilities and the listener's background.[2,3,4] Despite this theoretical framework, the complex interplay between speaker, listener, and L2 intelligibility remains relatively underexplored. Advances in machine learning, notably Deep Neural Networks (DNNs), have significantly improved the automated recognition of even L2 speech.[5] However, few studies in engineering have explored the pedagogical potential of using such tools in a classroom setting—a critical gap considering that individual learners are known to respond differently to identical instructional treatments (i.e., aptitude-treatment interaction).[6] The overarching aim of this proposal is to disentangle the roles of the speaker and the listener during real-world L2 speech communication using DNN-based analyses. We will develop a fully automated system to assess L2 speech based on different types of listeners (strict vs. lenient), and that can provide automated feedback on specific areas of language in need of improvement (phonology, fluency, and lexicogrammar). The research questions and accompanying hypotheses are as follows:

1. How do listeners with diverse backgrounds (e.g., native vs. non-native; linguistically trained vs. novice; positive vs. negative attitude towards foreign accents) evaluate the overall intelligibility of L2 speech?
2. In what ways does an automated feedback system support L2 speech development for learners with varied aptitude, motivation, and belief profiles?

Regarding RQ1, we intend to test the hypothesis that listeners with more experience with foreign-accented speech provide more lenient ratings of L2 speech.[7] We further aim to show via DNN-based analyses that these listeners' comprehension is linked to their ability to allocate equal attention to various linguistic elements in accented speech, thereby maximising their capacity to grasp the overall intended message.[8] Regarding RQ2, it has been posited that while automated feedback generally motivates L2 learners and leads to increased practice and gains, certain types of feedback are more effective for learners with specific individual difference profiles.[6] For example, learners with more precise auditory processing and superior working memory are likely to benefit from feedback provided by strict listeners.[9] Furthermore, metalinguistic feedback appears to be especially beneficial for learners whose motivation profiles are tied to form-oriented L2 speech learning.[10]

Significance

The theoretical contribution of the project will be a computational model of the bidirectional account of speech comprehension. The latest DNN-based machine learning paradigm will be used to disentangle the complex mechanisms underlying how different types of listeners (lenient, strict) use different types of speech information (segmentals, prosody, fluency, lexicogrammar) while judging the intelligibility and comprehensibility of L2 speech. A crucial innovation of the study is the use of posteriorgrams (rather than traditional spectrograms) to supress extra-linguistic factors typical of free speech (e.g., age, gender). Our pilot study (Saito, Minematsu et al., 2023, *SSLA*)[11] has validated this approach and found automated comprehensibility assessments of L2 speech to be more accurate than previous methods.

Building on this pilot study, the current project takes an initial step towards not only refining the assessment system with larger samples, but also exploring the question of how to integrate and optimise the use of this tool in language classrooms.

Method

In Phase 1 of the project, we will compile a speaker-listener corpus necessary for the development of the automated feedback system. Using the same methods outlined in our pilot study,[11] a total of 500 adult Japanese speakers of English with diverse proficiency levels will be recruited in Japan (classroom learners) and London, UK (immersion learners). Their spontaneous speech will be elicited using a range of speaking tasks. A total of 200 listeners with varied backgrounds (e.g., natives vs. non-natives, foreign-accent familiarity and attitude) will be recruited from the Greater London Area. Following well-established procedures from previous studies,[12] the listeners will listen to and assess the speech samples for ease of understanding (comprehensibility) on a 9-point scale, and then transcribe what they heard (intelligibility). They will be divided into lenient vs. strict listeners based on the results of cluster and Rasch analyses of their rating scores. DNN-based analyses will then simulate listeners' ratings of intelligibility and comprehensibility based on the segmental, prosodic, fluency, and lexicogrammar characteristics of the speech. Using an algorithm obtained from the corpus data, we will develop a fully-automated system that can provide an estimated degree of intelligibility and comprehensibility based on listener type (lenient, strict), and provide feedback on specific areas in need of improvement, including segmentals (consonants, vowels), prosody (stress, intonation), fluency (speed), and lexicogrammar (accuracy, complexity).[13]

In Phase 2, we will conduct a semester-long study with 200 Japanese university students to investigate the pedagogic potential of using the tool in L2 English speech training. Participants will first be randomly allocated into seven groups (see Table 1). All groups will commit to a minimum of 15 minutes of practice per day using the Gorilla platform.[14] The self-paced curriculum, informed by our previous work[15] and that of others,[16] will include monologue tasks and self-assessments of intelligibility. During the training, the training-only group will not receive any feedback, while the remaining participants will be divided into six distinct feedback groups based on listener factors (lenient, strict)[7] and metalinguistic information.[17] Groups 1 and 2 will receive feedback based on lenient listeners' ratings, while Groups 3 and 4 will receive feedback based on strict listeners' ratings. Groups 5 and 6 will initially receive feedback from lenient listeners, and gradually transition to strict listeners. Groups 1, 3, and 5 will be provided with metalinguistic feedback on specific areas requiring improvement, and will receive follow-up training based on their specific shortcomings in pronunciation, fluency, and lexicogrammar. In contrast, Groups 2, 4, and 6 will not receive such targeted information, and will be free to choose training areas which match their own interests and preferences.

Table 1 Summary of Content Across the Six Feedback Groups and the Training-Only Group

	Listener leniency	Metalinguistic information
Feedback 1 (n = 30)	Lenient Listeners	Yes
Feedback 2 (n = 30)	Lenient Listeners	No
Feedback 3 (n = 30)	Strict Listeners	Yes
Feedback 4 (n = 30)	Strict Listeners	No
Feedback 5 (n = 30)	Lenient to Strict Listeners	Yes
Feedback 6 (n = 30)	Lenient to Strict Listeners	No
Training-only (n = 20)	No Feedback Provided	N/A

Analyses

A series of mixed-effects regression analyses will be conducted to evaluate the extent to which participants' practice, as well as the comprehensibility and intelligibility of their L2 English speech (recorded during each session), change over time. To discern which learners are most responsive to various types of feedback (± listener leniency, ± metalinguistic information), their improvement scores in comprehensibility and intelligibility will be correlated with their volume of practice and individual differences in aptitude, motivation, and beliefs about language learning. These latter variables will be measured by a series of tasks assessing auditory processing,[18] working memory,[19] motivation,[20] and beliefs.[21]

Dissemination

We will submit papers to journals in the fields of education (e.g., *Language Learning*) and speech science (e.g., *Speech Communication*). To engage practitioners, we will host public symposia at University College London and the University of Tokyo, livestreaming and archiving them on our YouTube channel (2k+ subscribers, 100k+ views). Based on the project's findings and feedback from participants and peer review, we will update our automated speech training platform, and make it freely available to the public. The revised materials will be coded in JavaScript/HTML and uploaded to our L2 Speech Tools platform (http://sla-speech-tools.com/), where we have already shared extensive pronunciation resources (500k+ views since 2012).

1. Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.
2. Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707-729.
3. Trofimovich, P. (2016). Interactive alignment: A teaching-friendly view of second language pronunciation learning. *Language Teaching, 49*, 411-422.
4. Witteman, M. J. et al. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics, 75*, 537-556.
5. Wang, D., Wang, X., & Lv, S. (2019). An overview of end-to-end automatic speech recognition. *Symmetry, 11*, 1018.
6. DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189-200.
7. Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review, 64*, 459-489.
8. Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*, 866-900.
9. Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*, 634-654.
10. Yoshida, R. (2008). Teachers' choice and learners' preference of corrective feedback types. *Language awareness*, *17*, 78-93.
11. Saito, K., Minematsu, N. et al. (2023). Automated assessment of second language comprehensibility: Review, training, validation, and generalization studies. *Studies in Second Language Acquisition, 45,* 234-263.
12. Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company.
13. Saito, K. et al. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439-462.
14. Anwyl-Irvine, A. L. et al. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*, 388-407.
15. Saito, K., & Akiyama, Y. (2017). Video-based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language learning, 67*, 43-74.
16. Tsunemoto, A. et al. (2022). Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. *Foreign Language Annals, 55*, 135-154.
17. Ellis, R. et al. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339-368.
18. Saito, K., & Tierney, A. (in press). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S027226312200047X
19. Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
20. Saito, K. et al. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, *68*, 709-743.
21. Horwitz, E. K. (1999). Cultural and situational influences on foreign language learners' beliefs about language learning: A review of BALLI studies. *System*, *27*, 557-576.