FLUENCY BENCHMARKS AND IMPACTS OF PRACTICE WITH INSTANTANEOUS ASSESSMENT ON INTERNATIONAL TEACHING ASSISTANTS' SPEECH RATE AND PAUSE UNITS

Kevin Hirschi, Northern Arizona University
Okim Kang, Northern Arizona University
John Hansen, University of Texas at Dallas
Stephen D. Looney, Pennsylvania State University

This paper reports on the development of academic-specific L2 target benchmarks and introduces a web-based instantaneous assessment tool to provide second language (L2) English learners with quantitative results of their utterance speech rate and pauses. Thirty highly intelligible L2 speakers of English were recorded reading a paragraph aloud and responding to an explanation-based unconstrained speaking prompt to provide benchmarks for learners. Then, six pre-service International Teaching Assistants (ITAs) completed a Moodle course on pause units and speech rate with practice tasks that use the embedded instantaneous assessment tool. Speech captured at a pre- and post-tests was analyzed acoustically and evaluated by six trained human raters for appropriate speech rate, pause units, accentedness, and intelligibility. The results indicate that participants were able to apply the practice to their use of pause units and converge on the target speech rate, and that the changes resulted in trends of increased intelligibility but not accentedness. The paper gives insights into employing automated fluency analysis tools and digital pedagogical content for L2 learners, offers additional evidence of the relationship between suprasegmental features and listener perceptions, and provides support for the use of instantaneous assessment tools with L2 learners.

Cite as: Hirschi, K., Kang, O., Hansen, J., & Looney, S. D. (2023). Fluency benchmarks and impacts of practice with instantaneous assessment on International Teaching Assistants' speech rate and pause units. In R. I. Thomson, T. M. Derwing, J. M. Levis, & K. Hiebert (Eds.), *Proceedings of the 13th Pronunciation in Second Language Learning and Teaching Conference,* held June 2022 at Brock University, St. Catharines, ON.

INTRODUCTION

International Teaching Assistants (ITAs) at North American universities are often perceived to be problematic by their students due to differences in their speech relative to local speech norms (Inglis, 1993). As such, numerous institutions have created English skills training courses for effective communication, some of which employ the *Intelligibility Principle*. This approach provides a research-informed pedagogical roadmap for pronunciation training in which second language (L2) speech can be considered in terms of accentedness (i.e., the extent to which speech sounds different from a listener's speech community) and intelligibility (i.e., the extent to which the message is understood), the latter of which should be prioritized (Levis, 2005). The goal of pronunciation instruction that adheres to the *Intelligibility Principle* is therefore to promote intelligibility so that a wide variety of listeners can understand ITAs. Two important prosodic components of successful communication are speech rate (Kang et al., 2018; Munro & Derwing, 1998, 2001; Trofimovich & Baker, 2006) and pause units (Kahng, 2018; Tavakoli, 2011; Tyler et al., 1988).

Speech Rate and Pause Units

L2 speakers' speech rate, as measured by the number of syllables per second, is related to intelligibility (Kang et al., 2018). Studies of naïve listeners have found that speech rate increases are also related to decreases in perceptions of accentedness in English by Korean (Trofimovich & Baker, 2006) and Mandarin first language (L1) (Munro & Derwing, 1998) speakers. However, Munro and Derwing (2001) suggest that the relationship between speech rate and perception of L2 speech is curvilinear. That is, a rate that is too high or too low may work to undermine listener comprehension. Further exploration of this relationship across a wide variety of L1 backgrounds and proficiency levels has yet to confirm the acquisition of speech rate(s) that optimize intelligibility, especially considering the differences in speech rate found in L2 productions of dialogic vs. monologic speech (Tavakoli, 2016) and in relation to task familiarity (Lambert et al., 2017).

Pause units, or chunks of spoken discourse that are delimited by pauses and align with clause or phrase boundaries, are commonly produced by L2 learners. However, L2 learners may require processing time before the completion of a clause, resulting in mid-clausal pauses (Tavakoli, 2011). These unexpected pause patterns negatively impact perceptions of fluency (Kahng, 2018) and ITA effectiveness (Tyler et al., 1988). Although pedagogical materials promote explicit instruction and practice on pause units (Celce-Murcia et al., 2010; Grant, 2016), the impact of such training and its learnability remains unclear.

The Present Study

The present study seeks to further understand the relationship between speech rate and pause units and perceptions of ITA speech. It first reports on the development of benchmarks, then describes the online training effects of practice with instantaneous assessment on speech rate and pause features and on human listeners' perceptual judgments. It is guided by the following two questions:

RQ1. What are the effects of online training on the use of ITAs' speech rate and pauses?

RQ2. What are the impacts of online training on ITAs' intelligibility and accentedness?

METHODS

Participants

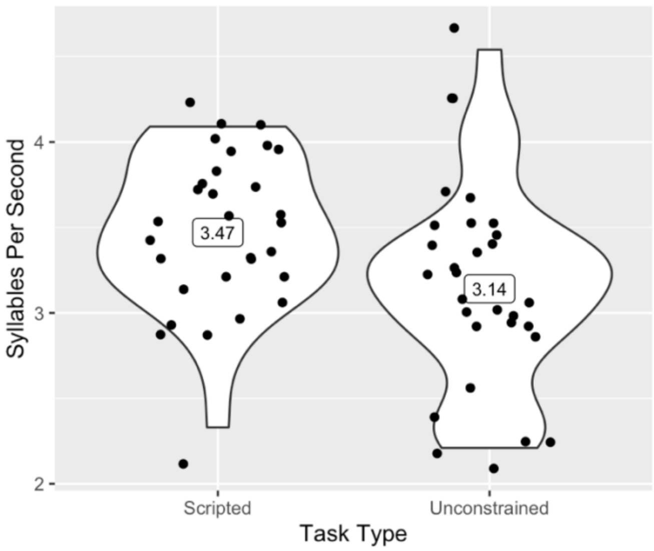
Prior to primary data collection, 30 highly intelligible L2 speakers with post-secondary teaching experience were recruited to provide benchmark speech rates and pause units. There were more females (n = 20) than males (n = 10) and their median age was 34 years old. Their L1 background were varied, including Spanish (n = 11), Chinese (n = 7), Russian (n = 3), Turkish, Korean, and others (n < 2 per group). For the intervention, six pre-service ITAs were recruited from ITA speaking skills training courses and ITA networks. Five were female and one participant was male. Their median age was 25 years old; five learners reported Chinese L1 backgrounds and one was L1 Spanish. For ratings of accentedness, intelligibility, and appropriateness of speech rate and pause units, six linguist raters (3 males and 3 females) were trained prior to rating ITA speech samples. Raters had a background in L2 phonetics/phonology and none reported abnormal hearing. Their median age was 36. Five were L1 English speakers from North America and one was from an L1 Spanish background.

Speech Rate and Pause Unit Targets

To inform target speech rate and pause unit scores from model speakers with similar backgrounds, recordings from 30 highly intelligible speakers were collected in an online spoken survey which contained eight tasks, one of which was scripted; the others were open-ended situation-based prompts. The present study uses only two of these tasks: the scripted paragraph read-aloud from Celce-Murcia *et al.* (2010) and an unscripted explanation task that models a classroom lecture. The 60 samples (i.e., two per speaker) that were on average 4.64 minutes in duration (SD = 65 sec.) were subjected to syllable rate and pause analysis using a Praat script by de Jong and Wempe (2009) with the silence threshold set to -25 dB, minimum pause duration set to 0.3 seconds, and minimum dip between peaks set to 2 dB. The Praat script results were compared to phonetician analysis of 30-second samples of 30 samples, resulting in a high correlation for speech rate ($r = 0.84$) and number of pauses ($r = 0.88$).

The speech rates in syllables per second of the highly intelligible model speakers were higher for the scripted task ($M = 3.47$, $SD = 0.43$) than the unscripted task ($M = 3.14$, $SD = 0.57$). The results were subjected to a paired samples *t*-test, indicating that there was a significant difference between task types, $t(53.70) = 2.52$, $p = .015$, $d = 0.66$. Therefore, two target speech rates are used as a benchmark for analysis in this paper (*scripted* = 3.47 syl/sec, *unconstrained* = 3.14 syl/sec). See Figure 1 for violin plots of highly intelligible speech rates.

Figure 1.
Violin plots of highly intelligible speaker speech rates on scripted and unconstrained tasks



Pause units of the utterances by the highly intelligible group were also analyzed for alignment with clause boundaries. Pauses longer than 0.3 seconds were tabulated and those within clauses were marked. A percentage of pauses was calculated by dividing the pauses at clause boundaries by the total number of pauses following De Jong (2016) but disregarding the use of filled pauses in order to mirror the automated fluency assessment tool which cannot detect filled pauses. The resulting pause unit alignment results were 96.1% for the scripted task and 83.7% for the unconstrained task.

Pre-Test and Post-Test

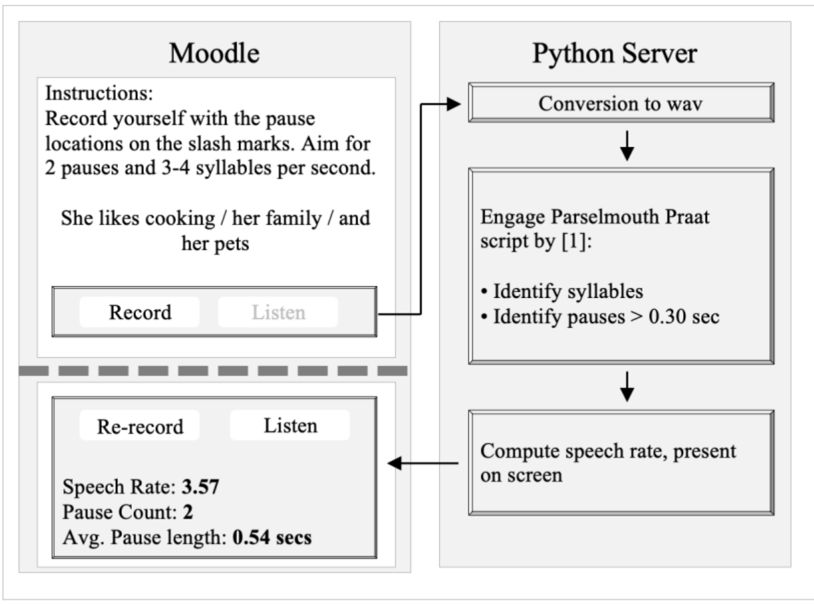
The ITA participants completed parallel versions of the speaking tasks on a pre-test and post-test with one scripted task and two unconstrained speech tasks, resulting in 5-7 minutes of speech captured at each point in time. The scripted task was a paragraph from (Celce-Murcia et al., 2010) which was chosen because it contains a wide variety of phonological features and exhibits the discourse tendencies of explanatory speech. The unscripted task asked the ITAs to explain a topic or concept, a common task of ITAs in classroom contexts. The ITA speaking tasks were the same as those given to the highly intelligible speaker group.

Online Course and Assessment System

The content of the lessons relied on materials for ITA and teacher training that target intermediate to advanced learners in the North American contexts (Celce-Murcia et al., 2010; Grant, 2016). In the lessons, learning tasks were presented both visually and audibly with highly intelligible speaker models providing various speech rates (e.g., below 2.00 syl/sec or above 5.00 syl/sec) and positive and negative examples of different pause locations (e.g., at the end of a clause, within a noun phrase) of the same target utterance. Learners first listened and chose the best option before proceeding to a practice page with a scripted or unconstrained task. Instantaneous assessment of speech rate and pauses was provided as feedback.

A speech rate feedback system was developed that captured learner speech using a JavaScript audio recorder that triggered the syllable identification Praat script (De Jong & Wempe, 2009). The speech rate feedback website is available to the public at the NAU Applied Linguistics Speech Lab website (https://sites.google.com/nau.edu/applied-ling-speech-lab-nau/projects/osim-project/speech-rate-analyzer). Once a recording is submitted for analysis, the website returns a speech rate value in syllables per second and pause count. A schema of the Moodle integration is depicted in Figure 2.

Figure 2.
Schematic diagram of instantaneous speech rate and pause assessment system



Procedures

Two pilot participants completed the lessons and provided feedback on the instructions, stimuli, and overall perceptions of the intervention tools. Revisions were made to increase saliency in written instructions and several stimuli were changed so that the linguistic content did not interfere with prosodic practice. The six ITAs interacted with the course over the period of three weeks. The lessons were short (< 15 minutes each) but were released on a weekly basis to encourage participants to revisit prior lessons and practice. All pieces of the study were contained within Moodle and were otherwise self-paced. A microphone check was required before beginning the pre-test and an introduction page provided navigation instructions.

Human Ratings

Once all participants completed the course, the audio files collected from the pre- and post-tests were reviewed manually for quality and completeness. Then they were shortened to 30 seconds and loaded into an online rating form that randomized the task type and data collection timepoint for each speech sample. The six human raters completed a training session with a member of the research team that included an overview of the project and target features. It took the raters approximately an hour to evaluate the 24 samples. The online rating form contained 100-point visual analogue scales (VAS) for intelligibility (% words understood, similar to Kang et al., 2018); accentedness (similar to Hirschi et al., 2020); and appropriateness of speech rate and pause units. Intra-class correlation was used to assess reliability for each dependent variable, indicating adequate rater consistency, for ratings of thought groups and speech rate appropriateness, ICC(3,k) = .84 and .82, respectively, as well as intelligibility and accentedness, ICC(3,k) = .81 and .76, respectively.

Analysis

For rating results, data were aggregated from the online rating system and subjected to Linear Mixed Effects Models (LMEMs) using the *lme4* (Bates et al., 2014), *lmerTest* (Kuznetsova et al., 2017), and *emmeans* (Lenth et al., 2022) packages in *R* (R Core Team, 2019). The models fit the dependent variables (e.g., appropriate speech rate, pause units, accentedness, intelligibility) to the fixed effects of time by task type and random effects of rater and speaker. LMEMs were chosen for their robustness with small sample sizes as well as the abilities to handle non-normal data and multiple observations per participant (Schielzeth et al., 2020).

For acoustic analyses, pre- and post-test data were subjected to the same Praat script (De Jong & Wempe, 2009) for speech rate values. Some ITAs produced utterances above 5.00 syl/sec on the pre-test. Therefore, a difference score was calculated for each observation by subtracting the benchmark speech rate from the observed speech rate and applying the absolute value function. In this case, a score of 0 is the benchmark score and a score of 1 is either one syllable per second higher or lower than the benchmark rate. Pause units were tabulated manually following the same procedures as the highly intelligible speakers.

RESULTS

Six human raters provided scalar judgments of accentedness, intelligibility, appropriateness of speech rate and pause units for 24 speech files, resulting in 576 data points. Results are presented in the order of the research question.

Intervention Effects on Speech Rate and Pause Units

Scripted and unscripted speech are visualized in Figure 3. The visualization indicates positive trends of speech rate and pause units for both types of speech. The LMEM results confirm a positive upward human rating score for pause units (t = 1.84, p < .001) and speech rate (t = 2.50, p = .014) from the pre-test to the post-test. No effects for task type were found in either model, indicating that gains were detected for both scripted and unscripted speech. Furthermore, interactions between task type (scripted and unscripted) and time (pre-test and posttest) were not significant. See Table 1 for estimated marginal means derived from the LMEM and statistical summary results in Appendix Table 4. Effect sizes were also calculated. The speech rate model was able to explain slightly more variance ($R^2_{mar.}$ = 0.07, $R^2_{cond.}$ = 0.43) than the pause unit model ($R^2_{mar.}$ = 0.05, R^2_{cond} = 0.48).

Figure 3.
Marginal means plot of human ratings of speech rate and pause units from pre- to post-test.

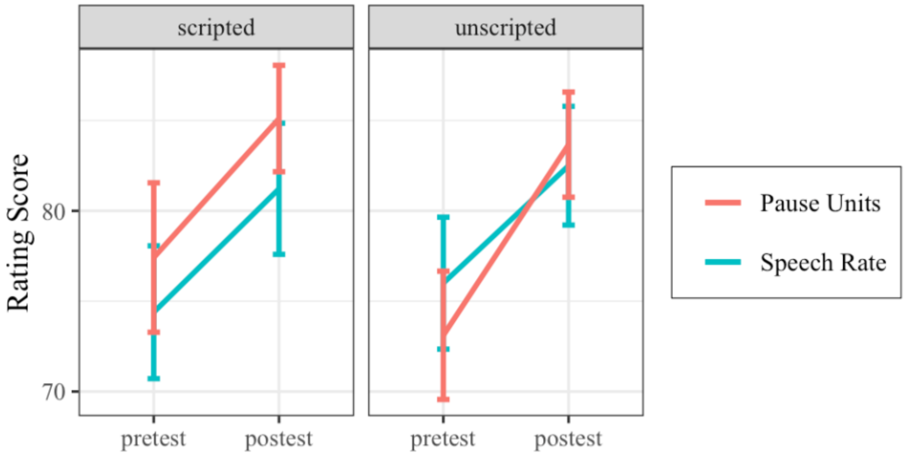


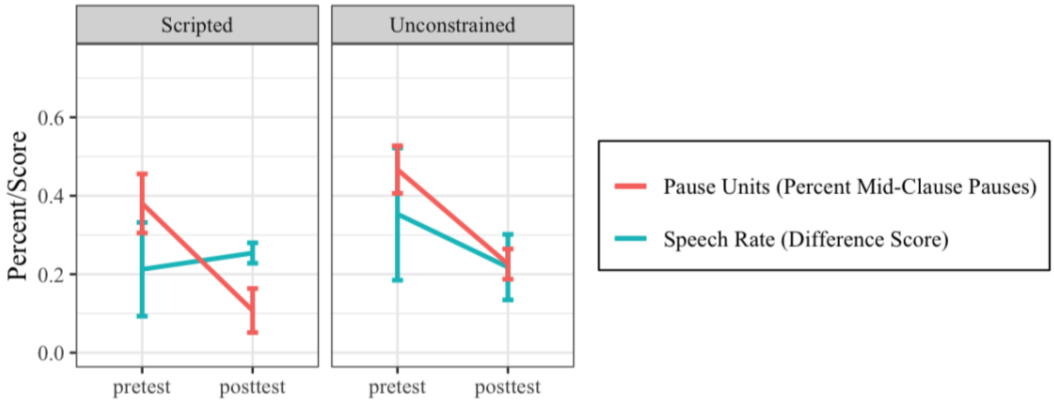
Table 1.
Estimated Marginal Means for Intervention Effects on Speech Rate and Pause Units (Human Raters) with Tukey HSD adjustments

	Scripted		Unscripted	
	M	*95% CI [LL, UL]*	*M*	*95% CI [LL, UL]*
		Speech Rate		
Pretest	74.39	[59.00, 89.78]	76.00	[60.61, 91.39]
Posttest	81.22	[65.83, 96.61]	82.50	[67.11, 97.89]
		Pause Units		
Pretest	77.42	[63.32, 91.52]	73.11	[59.01, 87.21]
Posttest	85.11	[71.01, 99.21]	83.67	[69.57, 97.77]

Note. LMEMs with random intercepts assume equal variance. Therefore, the standard error is equivalent across all cells for each construct and was 6.42 for Pause Units and 6.80 for Speech Rate.

To confirm the effects of the training detected by the human raters, acoustic analysis results of speech rate scores (i.e., the difference from the benchmark score to the utterance score) and pause unit score (the percentage of pauses placed mid-clause) were visualized using marginal means plots in Figure 4. For both of these values, a lower score is more target-like. See Table 2 for estimated marginal means and Appendix Table 5 for LMEM summary results.

Figure 4.
Marginal means plot of speech rate scores and pause units percentage from pre- to post-test (acoustic analysis, lower is better)



LMEM results on the acoustic measurements diverged from the human rater results in that no changes in speech rate score were detected from Time 1 to Time 2 (t = -2.06, p = .058). However, the variance around the speech rate decreased in time 2, indicating a convergence towards consistency even if the target rate was not achieved. The pause unit scores were significant from time 1 to time 2 (t = -8.28, p < .001) and there was a significant difference between the two task types (t = -3.22, p = .005), indicating that scripted speech was easier for ITAs to correctly place pauses. The effect sizes of the models diverged from the human perceptions in that the speech rate score model was explained less variance ($R^2_{mar.}$ = 0.11, $R^2_{cond.}$ = 0.45) than the pause units model ($R^2_{mar.}$ = 0.48, $R^2_{cond.}$ = 0.86).

Table 2.
Estimated Marginal Means for Intervention Effects on Speech Rate and Pause Units (acoustic analysis) with Tukey HSD adjustments

	Scripted		Unscripted	
	M	*95% CI [LL, UL]*	*M*	*95% CI [LL, UL]*
Speech Rate				
Pretest	0.69	[0.15, 1.21]	0.84	[0.31, 1.36]
Posttest	0.43	[-0.10, 0.95]	0.31	[-0.22, 0.84]
Pause Units				
Pretest	0.38	[0.24, 0.52]	0.47	[0.33, 0.60]
Posttest	0.10	[-0.03, 0.24]	0.22	[0.08, 0.36]

Note. LMEMs with random intercepts assume equal variance. Therefore, the standard error is equivalent across all cells for each construct and was 0.06 for Pause Units and 0.25 for Speech Rate.

Impacts on Accentedness and Intelligibility

Ratings on accentedness and intelligibility were similarly compiled and subjected to LMEM analysis. The results indicated that there was a positive trend of increased intelligibility for both scripted and unscripted tasks but mixed trends for accentedness. See Figure 5.

The LMEM results confirm that only a significant difference was detected on intelligibility (t = 2.40, p = .018), but it was detected in both scripted and unscripted tasks. No effects were significant in the accentedness model. See Table 3 for estimated marginal means and Appendix Table 6 for LMEM summary results. The effect size calculations indicate the models explained minimal variance amongst the fixed effects for intelligibility ($R^2_{mar.}$ = 0.02, $R^2_{cond.}$ = 0.59) and accentedness ($R^2_{mar.}$ = 0.01, $R^2_{cond.}$ = 0.55).

Table 3.
Estimated Marginal Means for human ratings of intelligibility and accentedness with Tukey HSD adjustments

	Scripted		Unscripted	
	M	*95% CI [LL, UL]*	*M*	*95% CI [LL, UL]*
		Intelligibility		
Pretest	84.34	[72.44, 96.24]	81.67	[69.77, 93.56]
Posttest	86.53	[74.63, 98.44]	88.03	[76.12, 99.93]
		Accentedness		
Pretest	46.56	[28.42, 64.69]	41.81	[23.67, 59.94]
Posttest	43.73	[25.58, 61.89]	43.87	[25.72, 62.03]

Note. LMEMs with random intercepts assume equal variance. Therefore, the standard error is equivalent across all cells for each construct and was 5.41 for Intelligibility and 8.26 for Accentedness.

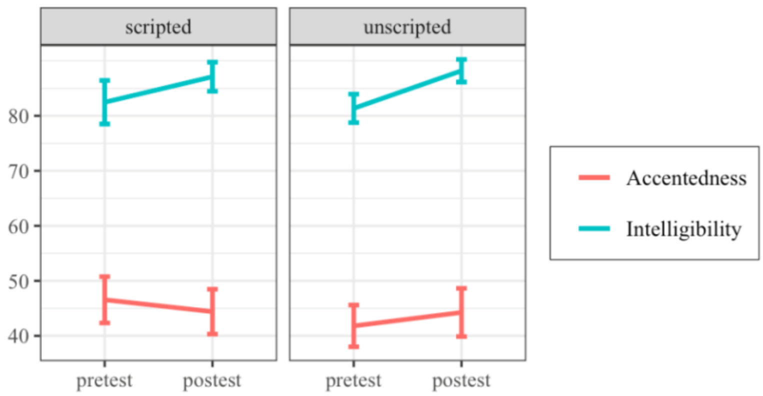


Figure 5: Marginal means plot of human ratings of accentedness and intelligibility from pre- to post-test

DISCUSSION

Taken together, the results of the current study indicate that the intervention with instantaneous assessment of speech rate and pauses is effective for training L2 speakers to place pauses in appropriate locations, which in turn, results in a small but detectable increase in intelligibility. While the human raters were able to detect differences in speech rate and pause unit production from pre- to post-test, acoustic analysis only detected differences in pause unit production. This discrepancy between human perceptions and acoustic analysis is not uncommon (Kang, 2010).

The findings of the highly intelligible benchmark rates are also of interest and a desirable goal (Munro & Derwing, 1998). While L1 speakers may use higher rates than those found in the present study, highly-intelligible L2 speaker benchmark rates remain within the 3-3.5 syllable per second range and vary between scripted and unconstrained speech. The study also confirms previous research on the relationship between speech rate and listener perceptions (Kang, 2010), and extends the relationship to pause units.

However, this study is limited in that generalizations cannot be made from six participants and without a control group. Similarly, the use of six raters for perceptual variables such as accentedness, intelligibility, and appropriateness of prosody may also inhibit the robustness of the findings. Additionally, the differential impact of a slower or faster speech rate on intelligibility amongst ITAs is not fully understood, and the incorporation of empirically informed ranges based on larger sample sizes is necessary for future large-scale suprasegmental interventions. Furthermore, the use of a visual analogue scale to measure intelligibility in this study has received criticism from reviewers, as it diverges from the established transcription-based intelligibility scoring technique and may include perceptions of L2 speech that are not entirely associated with intelligibility. Finally, task repetition effects may explain some of the increases from the pre- to post-test despite the three-week gap (Lambert et al., 2017).

CONCLUSION

This study explored the development of benchmarks and an instantaneous assessment tool for L2 practice of speech rate and pause units. It found that training is likely effective in increasing pause placement accuracy and adjusting speech rate for both scripted and unconstrainted tasks. The training also resulted in a positive impact of intelligibility. Findings are limited with a small sample size and a short training period. However, they offer important implications regarding the importance and learnability of suprasegmental features to the fields of speech production and perception, language acquisition, and curriculum development and assessment.

ACKNOWLEDGEMENTS

This work is supported by NSF EAGER 2140414 funding “Second Language Speech Production: Formulation of Objective Speech Intelligibility Measures and Learner-Specific Feedback.”

ABOUT THE AUTHORS

Kevin Hirschi is a doctoral candidate in the Applied Linguistics program at Northern Arizona University. His interests include second language pronunciation, corpus linguistics, and mobile-assisted pronunciation technology.

Northern Arizona University
English Department
705 S Beaver St, Flagstaff, AZ 86011
KevinHirschi@nau.edu

Okim Kang is a Professor in the Applied Linguistics Program at Northern Arizona University, Flagstaff, AZ, USA. Her research interests are speech production and perception, L2

pronunciation and intelligibility, L2 oral assessment and testing, automated scoring and speech recognition, World Englishes, and language attitude.

John Hansen serves as Associate Dean for Research, and Professor of Electrical Engineering, Speech & Hearing Sciences at Univ. of Texas at Dallas. He oversees the Center for Robust Speech Systems and serves as ISCA President. He has authored/co-authored 752 papers in the field of speech processing and language technology.

Stephen D. Looney is a Teaching Professor of Applied Linguistics and Director of the International Teaching Assistant Program at the Pennsylvania State University. His research employs qualitative and mixed methods approaches to investigate classroom interaction, assessment of L2 interaction, and L2 pronunciation teaching and learning.

REFERENCES

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-31) [R Package]*.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: A course book and reference guide*. Cambridge: Cambridge University Press.

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 113–132.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Grant, L. (2016). *Well said: Pronunciation for clear communication* (Fourth). Heinle & Heinle.

Hirschi, K., Kang, O., Cucchiarini, C., Hansen, J., Evanini, K., & Strik, H. (2020). Mobile-Assisted Prosody Training for Limited English Proficiency: Learner Background and Speech Learning Pattern. *Proc. Interspeech 2020*, 4452–4456.

Inglis, M. (1993). The communicator style measure applied to nonnative speaking teaching assistants. *International Journal of Intercultural Relations*, *17*(1), 89–105. https://doi.org/10.1016/0147-1767(93)90014-Y

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569–591.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. System, 38(2), 301–315. doi:10.1016/j.system.2010.01.005

Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning*, *68*(1), 115–146.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167–196.

Lenth, R. V., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). *Estimated Marginal Means, Aka Least-Squares Means (Version 1.8.2) [R Package]*. https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369-377.

Munro, M. J., & Derwing, T. M. (1998). The effects of speech rate on the comprehensibility of native and foreign accented speech. *Language Learning*, *48*(2), 159–182.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in Second Language Acquisition*, *23*(4), 451–468.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152.

Tavakoli, P. (2011). Pausing Patterns: Differences between L2 Learners and Native Speakers. *ELT Journal*, *65*(1), 71–79.

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 133–150.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30.

Tyler, A. E., Jefferies, A. A., & Davies, C. E. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teacher's spoken discourse. *World Englishes*, *7*(2), 101–110.

Appendix

Table 4

LMEM summary results of rater perceptions of speech rate and pause units

Speech Rate

Random Effects	Variance	*SD*		
Subject	131.85	11.48		
Rater	74.39	8.63		
Fixed Effects	Est.	*SE*	*t*	*p*
(intercept)	78.53	6.39	12.28	<.001***
Time2	4.71	1.88	2.50	.014***
TaskType2	-0.72	1.33	-0.54	.589
Time*TaskType	0.12	1.88	0.06	.950

Pause Units

Random Effects	Variance	*SD*		
Subject	188.33	13.72		
Rater	46.23	6.80		
Fixed Effects	Est.	*SE*	*t*	*p*
(intercept)	79.83	6.01	13.29	< .001***
Time2	6.45	1.84	3.50	<.001*
TaskType2	1.44	1.30	1.10	.272
Time*TaskType	-1.01	1.84	-0.55	.584

Table 5

LMEM summary results of acoustic analysis of speech rate and pause units

Speech Rate Score

Random Effects	Variance	*SD*		
Subject	0.14	0.38		
Fixed Effects	Est.	*SE*	*t*	*p*
(intercept)	0.56	0.18	3.13	<.026*
Time2	-0.28	0.14	-2.06	.058
TaskType2	-0.01	0.10	-0.10	.925
Time*TaskType	0.09	0.14	0.69	.502

Pause Unit Percentage

Random Effects	Variance	*SD*		
Subject	0.02	0.12		
Fixed Effects	Est.	*SE*	*t*	*p*
(intercept)	0.30	0.05	5.61	.002**
Time2	-0.18	0.02	-8.27	<.001***
TaskType2	-0.05	0.02	-3.29	.005**
Time*TaskType	-0.01	0.02	-0.52	.614

Table 6

LMEM Summary results of human ratings of intelligibility and accentedness

Intelligibility

Random Effects	Variance	*SD*		
Subject	85.39	9.24		
Rater	71.35	8.45		

Fixed Effects	Est.	*SE*	*t*	*p*
(intercept)	85.14	5.19	16.41	<.001***
Time2	3.02	1.26	2.40	.018*
TaskType2	0.29	0.89	0.33	.742
Time*TaskType	-1.47	1.26	-1.17	.245

Accentedness

Random Effects	Variance	*SD*		
Subject	157.60	12.56		
Rater	202.01	14.21		

Fixed Effects	Est.	*S*	*t*	*p*
(intercept)	43.99	7.87	9.23	<.001***
Time2	-0.27	2.04	-0.13	.896
TaskType2	1.15	1.44	0.80	.426
Time*TaskType	-1.73	2.04	-0.85	.398