

PRESENTATION/POSTER

NATIVE LISTENERS' EVALUATIONS OF PLEASANTNESS, FOREIGN ACCENT, COMPREHENSIBILITY, AND FLUENCY IN THE SPEECH OF ACCENTED TALKERS

[Jeun Lee](#), University of Wisconsin-Milwaukee

[Dong Jin Kim](#), University of Wisconsin-Milwaukee

[Hanyong Park](#), University of Wisconsin-Milwaukee

Compared to studies on accentedness, fluency, and comprehensibility, there are few studies on pleasantness in second language (L2) research. To address this gap, we investigated native English speakers' subjective evaluations of pleasantness, accentedness, comprehensibility, and fluency in the speech of Korean learners of English. Twenty-six raters made judgements on a 9-point Likert scale after listening to spontaneous speech samples from Korean learners of English. Results indicated that pleasantness ratings are predicted by all dimensions. In particular, we observed fluency as the best predictor of native listeners' perceived pleasantness, followed by comprehensibility and accentedness. Our findings suggest native speakers' (NSs) appraisals of L2 speech pleasantness is influenced by how fluent and comprehensible L2 speakers are.

INTRODUCTION

Listeners automatically evaluate many aspects of the speech of their interlocutors. The degree of foreign accent, comprehensibility, fluency, and intelligibility are some aspects that listeners continuously assess while conversing with non-native speakers (NNSs). Previous research has examined the relationships between these dimensions in NSs' perception of utterances by L2 learners (e.g., Derwing, Rossiter, Munro, & Thomson, 2004; Munro & Derwing, 1995a, 1999; Trofimovich & Isaacs, 2012). Pleasantness, listeners' subjective evaluation of their holistic conversational experience, may be another important dimension in L2 learning in that pleasant speech may result in successful encounters with NSs and this may increase learners' motivation to seek out more interactions. In spite of the potentially important role of pleasantness in L2 learning, there are few studies on pleasantness. In particular, questions regarding native listeners' reactions to L2 speakers' pleasantness and to what extent this dimension is related to other speech dimensions are mostly unanswered. Thus, we investigate these questions in the current study.

Research on accentedness, comprehensibility, and fluency

The overall conclusion about the relationships between accentedness, comprehensibility, and fluency is that these dimensions are associated with each other with varying degrees of strength. For example, Derwing, Rossiter, Munro, and Thomson (2004) found a stronger relationship between fluency and comprehensibility than between fluency and accentedness in NSs' ratings of Mandarin learners of English. Their results suggest that a strong foreign accent does not necessarily result in reduced fluency or comprehensibility. However, to the best of our knowledge, there is no research that attempts to examine how pleasantness constructs relationships with other

L2 speech aspects (accentedness, comprehensibility, and fluency). The major goal of the present study is to address this gap by asking native English listeners to judge the same set of stimuli on the above-mentioned four dimensions.

Research on pleasantness

Previous speech perception studies approached pleasantness as listeners' perceived (subjective) attitude toward some particular aspects of speech, such as accent, pronunciation, or voice. For instance, Giles (1970) studied listeners' perceived attitudes of the "aesthetic" content (pleasantness) of 13 different accents of native language (L1) presented both vocally and conceptually. Participants made attitude judgements on a 7-point scale (1 = extremely pleasant; 7 = extremely unpleasant). The results suggested that perceived attitudes vary for different accents, and non-linguistic factors such as age, sex, social class, and regional membership are important determinants of listeners' evaluations. Jakšič (2018) studied native Czech English as a Second Language (ESL) learners' judgements of six different varieties of English on comprehensibility, pleasantness, socioeconomic status, and model suitability. The instruction given for pleasantness ratings was "For me, the speaker's pronunciation sounds: 1 = very pleasantly [sic]; 7 = very unpleasantly [sic]" (p. 46). As in Giles (1970), listeners' judgements of pleasantness were different for several dialects.

Similar to the studies of the perceived pleasantness of speech produced by native speakers, Bouchard, Carranza, and Moffie (1977) investigated native listeners' perceived pleasantness of speech produced by L2 speakers while focusing on voices. In their study, native English listeners judged Spanish-English bilinguals' taped readings of an English passage on the likelihood of being a friend, eventual occupation, accentedness, pleasantness, and fluency using 7-point scales. Raters were told to assess each recording on the basis of voice cues alone, just as one might judge a person if the individual were talking on the telephone or speaking on the radio. The correlations among all five rating dimensions were statistically significant, and pleasantness showed the strongest positive correlation with fluency.

Along with Bouchard *et al.* (1977), which showed the close relationship between pleasantness and fluency, Derwing and Munro's (2009) study on preference suggests a possible relationship among pleasantness, comprehensibility, and fluency. They examined how L2 speech comprehensibility influences native English listeners' preference for interacting with Mandarin and Slavic learners of English. After listening to a pair of extemporaneously spoken L2 speech samples, the listeners chose the sample that they preferred. They also had an opportunity to write comments about each sample they heard. Overall, they preferred more comprehensible speech regardless of the speaker's L1. It is also worth mentioning that comprehensibility-related (e.g., easy to follow) and fluency-related (e.g., broken speech) comments were a large part of comments for the selected (preferred) voices (41%) and the non-selected (non-preferred) voices (51%). Although pleasantness in the current study and preference in Derwing and Munro's (2009) study are not identical concepts, these two have something in common; both of them would be likely to lead more interaction between L2 learners and NSs.

The current study

Unlike previous studies, in the current study on the relationships between pleasantness and accentedness, comprehensibility, and fluency, we focus on English native listeners' impressions of pleasantness in a holistic way rather than asking them to pay attention to certain features of speech. We took this approach to pleasantness because we were not confident that listeners would be able to evaluate a speech dimension and focus on one aspect of speech while ignoring others. Thus, we asked the listeners to simply rate the pleasantness of each L2 speech on a 9-point Likert scale.

This study is guided by the following research question: How is pleasantness related to accentedness, comprehensibility, and fluency? Based on previous research, we predict that pleasantness is more closely aligned with fluency and comprehensibility than accentedness.

METHODS

Talkers

Twenty-one L1-Korean speakers of L2-English (15 females, 6 males) produced the stimuli for monetary compensation. They were either university students or residents living in the Midwestern USA (mean age = 27.14, $SD = 6.21$). The average length of residence (LOR) of the L2 speakers was 2.82 years ($SD = 1.95$) and the average age of arrival (AOA) was 23.90 years ($SD = 5.77$). These production data were collected as part of a larger study conducted by Darcy, Park and Yang (2015).

Listeners

Twenty-six NSs of American English (18 females and 5 males) participated in the current study as raters for course credit. They were university students in the Midwest (mean age = 25.54 years, $SD = 7.66$).

Procedures

The L2 speakers participated in a narrative retelling task of a summary of the *North Wind and the Sun*. After reading the passage on a computer screen, they retold the story. The participants were unaware that they would be retelling the story when they read the passage. Recordings were made in a sound-attenuated room with a microphone at a sampling rate of 44.1 kHz. Excluding the first and the last sentences of the utterances of each speaker, two sentence-long tokens per speaker were selected as stimuli. These criteria for token selection were to avoid speakers' initial hesitation and disfluencies which might appear due to the sudden start of the retelling task and also prevent raters from being bored due to a lengthy experiment. Although sentence-long stimuli are not commonly used in studies on L2 fluency, we have decided to use somewhat short stimuli based on previous studies (e.g., Munro & Derwing, 2001) reporting that listeners can make reliable judgements on different aspects of L2 speech after listening to sentence-long stimuli. Another thing to consider is that our study was conducted with 42 speech samples from 21 non-native talkers and had two rating sessions (see Rating on page 171). If we followed the commonly used methods such as 30

seconds-length stimuli, the experiment would have taken almost 50 minutes, which is too long for a perceptual experiment (e.g., if we give 5 seconds to evaluate two speech dimensions for each speech sample, it takes 2,940 seconds to finish the experiment without any break: 42 tokens x (30 seconds + 5 seconds) x 2 sessions = 2,940 seconds or 49 minutes).

Filled pauses were discarded and unfilled pauses (i.e., silence) of over three seconds were modified by removing the part of pause after three seconds threshold. For example, if an unfilled pause was 3.04 seconds, only three seconds of pause remained, and the excessive 0.04 seconds were removed. Considering we used sentence-length stimuli, we were worried that too many filled pauses and extremely long unfilled pauses would becloud raters' evaluations of the L2 speech dimensions of our interest, result in a very skewed distribution in fluency ratings, and consequently make it difficult to investigate the relationships of perceptual dimensions including fluency. Although we manipulated two fluency characteristics of the samples regarding pauses, it should be noted that we considered only *uhs* and *ums* as filled pauses by following Lee (2018) and did not manipulate most characteristics of fluency, such as repetition, replacements, reformulations, hesitations, and false starts. Also, such a three seconds pause manipulation was processed only for two out of forty-two stimuli (approximately 4.76 of total stimuli). After our manipulations, we still observed a wide range of fluency ratings, suggesting that there were many fluency characteristics other than the two characteristics we manipulated. Furthermore, since our main interest is not fluency itself but the relationships among four L2 perceptual dimensions, we felt that it would be fair to examine those relationships as long as raters were given the same stimuli regardless of the manipulations of stimuli. After the editing process, the average length of selected tokens was 8.70 seconds. In total, forty-two tokens of L2-English speech from twenty-one native Korean speakers were used as stimuli in the rating session.

Rating

The experiment was conducted on a Praat platform (Boersma & Weenink, 2017) with high-quality headphones at the UWM Phonetics Lab. After completing the consent form and a language background questionnaire, the raters participated in a practice session to familiarize them with the rating task. The task consisted of two parts, with a minute of mid-session break. It took about thirty minutes to complete the rating task. In each session, the participants rated two dimensions on a 9-point Likert scale after listening to each token. For example, participants rated pleasantness and accentedness for all 42 tokens in the first session, then, after the break, they rated comprehensibility and fluency for the same set with a different randomization. All possible twenty-four combinations of rating dimensions (e.g., accentedness-comprehensibility, pleasantness-fluency) were considered and the stimuli were presented in randomized sequences. Following the definitions suggested by previous studies (e.g., Derwing & Munro, 1997; Kormos & Dénes, 2004; Munro & Derwing, 1995b; Trofimovich & Isaacs, 2012; Zetterholm & Abelin, 2017), the rating dimensions were defined as follows: Pleasantness—how pleasant or unpleasant your experience of listening to the sentence is (1 = very unpleasant, 9 = very pleasant); Accentedness—how different the speaker's accent is from standard American English (1 = very strong foreign accent, 9 = no foreign accent); Comprehensibility—how easy or difficult it is to understand the sentence (1 = impossible to understand, 9 = very easy to understand); Fluency—how fluent or disfluent the speaker is (1 = very disfluent, 9 = very fluent). These definitions were given to the raters before the practice and

rating sessions on a separate piece of paper and during the sessions on a computer screen. The participants were asked to use the full-scale range when making their judgements.

RESULTS

To answer the research questions regarding the relationships among dimensions, we ran a mixed-effects model with lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2017) after excluding outliers of the rating data based on the 3SD threshold (Kennedy & Trofimovich, 2008). Altogether, five instances in pleasantness, two in accentedness, one in comprehensibility, and six in fluency were excluded.

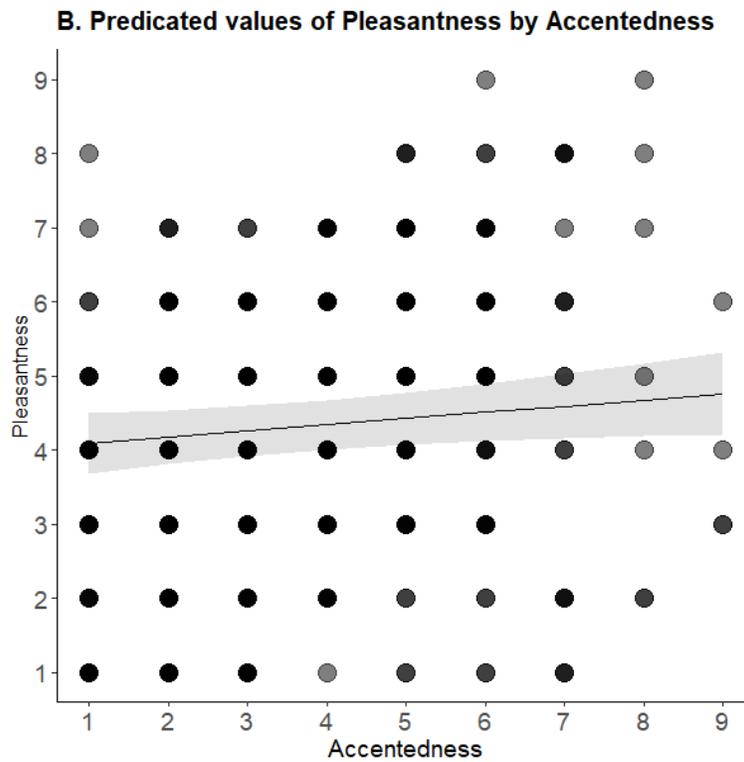
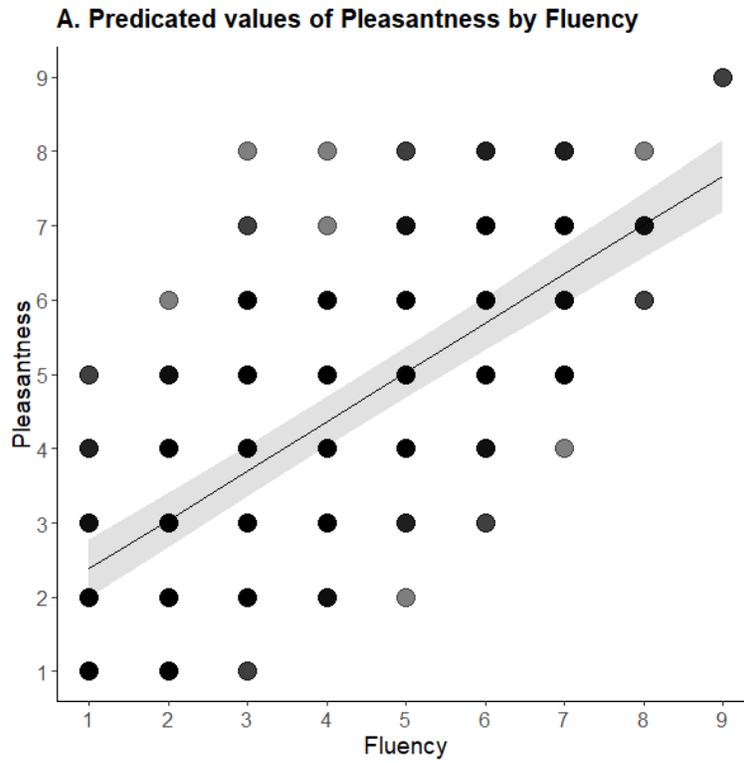
The relations among pleasantness, accentedness, comprehensibility, and fluency

Table 1 shows the linear mixed effects model of pleasantness ratings. The model contained accentedness, comprehensibility, and fluency as fixed effects. Intercepts for raters and stimuli were included as random effects. P-values and r^2 were calculated using lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) package and Nakagawa and Schielzeth approach (Nakagawa & Schielzeth, 2013) in r2glmm (Jaeger, 2017) package. The type 3 analysis of variance found that there were significant main effects of fluency (estimate = .661, $SE = .043$, $t = 18.51$, $F(1, 450.97) = 342.64$, $r^2 = .422$, $p < .001$) and comprehensibility (estimate = .091, $SE = .041$, $t = 2.217$, $F(1, 396.06) = 4.92$, $r^2 = .015$, $p = .027$), as well as a trend effect of accentedness (estimate = .084, $SE = .043$, $t = 1.939$, $F(1, 443.18) = 3.76$, $r^2 = .009$, $p = .053$). The percentages of variance explained by the variables were 42%, 1.5%, and 0.9% respectively. Overall, more fluent, more comprehensible, and less accented L2 speech was more likely to be evaluated as being more pleasant speech. Figure 1 summarizes the results.

Table 1

The output of linear mixed effects regression models of pleasantness with fluency, comprehensibility and accentedness

Predictors	pleasantness				
	Estimates	SE	t	Rsq	p
(Intercept)	.978	.228	4.286	.547	<.001
Fluency	.661	.036	18.51	.422	<.001
Comprehensibility	.091	.041	2.217	.015	.027
Accentedness	.084	.043	1.939	.009	.053
Observations	455				



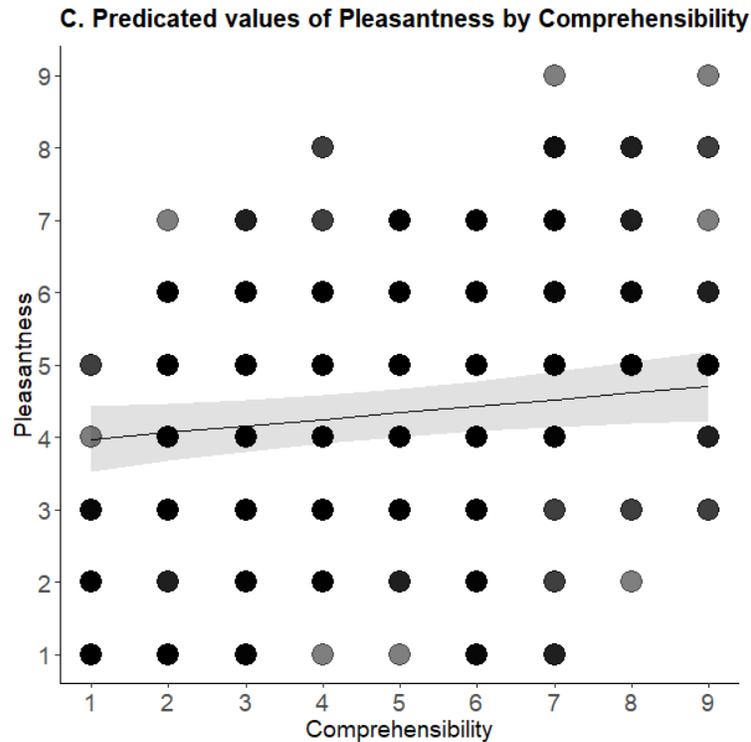


Figure 1. Prediction plots for pleasantness by fluency, accentedness, and comprehensibility. The lines show smoothed linear trends of the model-predicted effects. The shadings indicate 95% confidence-interval band. The darker dots, the more observations were made.

DISCUSSION AND CONCLUSIONS

This study investigated native listeners' evaluation of pleasantness, accentedness, comprehensibility, and fluency and their inter-relationships in L2 speech. Our study confirms previous findings that perceptual dimensions are connected to one another to varying degrees. Specifically, regarding pleasantness, our results showed that ratings of fluency seemed to be the strongest predictor of listeners' perceived pleasantness of L2 speech, followed by ratings of comprehensibility and accentedness. Our study provides support for Bouchard *et al.*'s (1977) findings. It is noteworthy that participants in their study were directed to focus on voice cues alone while rating Spanish-English bilinguals' taped readings of an English passage. It seems that in spite of listeners' special attention to bilinguals' voices, fluency still had a relatively large impact on pleasantness ratings. One possible explanation for this result is that the English passage reading stimuli in their study were quite long (forty-one words) from which listeners might use other cues, such as pronunciation errors or speech rate, along with voice cues to determine speakers' fluency levels. Adopting such long stimuli may have resulted in similar results in our study where no specific instruction was given regarding the basis of evaluation. Thus, together with Bouchard *et al.*'s (1977) findings, the results of the current study suggest that listeners' holistic impression of L2 speech pleasantness is mainly affected by how fluent L2 speakers sound.

Our predictions regarding the relationship between pleasantness, fluency, comprehensibility based on Derwing and Munro (2009) were also borne out in this study. Derwing and Munro (2009)

reported that English native listeners preferred to interact with L2 speakers who had more comprehensible L2 speech. The listeners also frequently commented on comprehensibility- and fluency-related features of L2 speech when explaining their preference. Consistent with their findings, NSs' pleasantness ratings in this study were mostly predicted by fluency and comprehensibility ratings. We also noted that fluency could explain a large portion (42%) of the variance in pleasantness ratings while comprehensibility explained only 1.5% of the variance in this study. These different contributions of fluency and comprehensibility to pleasantness ratings might come from the harsher judgements on fluency than on comprehensibility by listeners in the current study (Derwing, Munro, & Thomson, 2007; O'Brien, 2014). Listeners gave the highest ratings (indicating the most positive assessment. e.g., 9 = extremely fluent) to stimuli more often for comprehensibility than for fluency or pleasantness. Thus, the variance of comprehensibility ratings was larger than the ones of fluency and pleasantness, which showed more positive skewed patterns (i.e. more frequent ratings for lower scores, which means more negative assessments). These similar patterns of pleasantness and fluency might result in the closer relationship between these two dimensions compared to comprehensibility and accentedness. In sum, fluency is proposed as a best predictor of NSs' perceived pleasantness of L2 speech in this study, although the impacts of other dimensions still exist.

A question is in order regarding the strong relationship between pleasantness and fluency: why are these dimensions closely related in evaluating L2 speech? In a follow-up study (Kim, Lee, & Park, 2018), we explored whether fluency-related linguistic properties also affected pleasantness judgements. In this vein, we measured speech rate, repair fluency, and mean length of run (Derwing et al., 2004; Kormos & Dénes, 2004; Trofimovich & Isaacs, 2012). Our results indicated that pleasantness rating scores are significantly correlated with all measures. These results suggest that the strong relationship between pleasantness and fluency ratings in the current study may come from the common linguistic properties affecting both dimensions.

Pleasantness is an important aspect of L2 speech in that more pleasant speech may increase the amount of positive verbal and non-verbal feedback from NSs, since pleasant speech may increase NSs' desire to interact with L2 learners. As a consequence, L2 learners' motivation to communicate in their target language may increase by having successful encounters with NSs. The Douglas Fir Group (2016) stated that "For L2 learners ... the more they (L2 learners) experience emotionally and motivationally positive evaluations of their anticipated and real interactions, the more effort they will make to participate in them and affiliate with others" (p.28). Through this process, L2 learners are expected to build higher L2 confidence on the basis of positive L2 experiences. The importance of enhancing L2 confidence for successful L2 learning cannot be overstated (MacIntyre, 2007).

One remaining question for teachers is what might the positive pedagogical strategies for improving pleasantness be? Our study does not have a direct pedagogical component. However, based on our findings, we propose that utilizing existing pedagogies for fluency or comprehensibility improvement may be helpful.

ACKNOWLEDGMENTS

We thank Dr. Isabelle Darcy, who kindly allowed us to analyze the data for our study. We also thank John Levis, Charles Nagle, and Erin Todey and reviewers (Lucy Pickering and one anonymous reviewer) for their constructive feedback on our manuscript and the audience at PSLLT for their positive feedback on our presentation. Lastly, we acknowledge travel support from the UWM graduate school, Linguistics Department, to present our results at PSLLT.

ABOUT THE AUTHORS

Jieun Lee is a PhD student of Linguistics at the University of Wisconsin-Milwaukee (UWM). Her research interests concern L2 acquisition, phonetics & phonology, and L2 training methods.
Email: jieunlee@uwm.edu

Dong Jin Kim is a PhD student of Linguistics at UWM. His interests lie in phonetics, phonology, and prosody.
Email: kim252@uwm.edu

Dr. Hanyong Park is an associate professor at the Linguistic department at UWM. He is interested in L2 speech perception/production and effective L2 training methods.
Email: park27@uwm.edu

REFERENCES

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Bouchard R. E., Carranza, M. A., & Moffie, R. W. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals. *Language and Speech*, 20(3), 267-273.
- Browman, P., & Weenink, D. (2017). Praat: Doing phonetics by computer (Version: 6.0.43). Amsterdam, The Netherlands: University of Amsterdam. Retrieved from <http://www.praat.org/>
- Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63-72.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1-16.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359-380.
- Derwing, T. M., & Munro, M. J. (2009). Comprehensibility as a factor in listener interaction

- preferences: Implications for the workplace. *Canadian Modern Language Review*, 66(2), 181-202.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: judgments on different tasks. *Language Learning*, 54(4), 655–679.
- Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, 100(S1), 19–47.
- Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, 22(3), 211–227.
- Jaeger, B. (2017). r2glmm: computes r squared for mixed (multilevel) models. R package version 0.1.2. Retrieved from <https://CRAN.R-project.org/package=r2glmm>.
- Jakšič, J. (2018). *Comprehensibility and pleasantness of different varieties of English as judged by Czech listeners* (Doctoral dissertation, Univerzita Karlova). Retrieved from <https://dspace.cuni.cz/handle/20.500.11956/100028>
- Kim, D. J., Lee, J., & Park, H. (2018, October). Some candidates for evaluating fluency, accentedness, comprehensibility, and pleasantness of non-native speech. The 23rd Annual Mid-Continental Phonetics & Phonology Conference, Northwestern University, Evanston, IL.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164.
- 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.
- Lee, S. (2018). *Effective Planning in Real-Time Speaking Test Tasks*. Unpublished Doctoral dissertation, Michigan State University.
- MacIntyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal*, 91(4), 564-576.
- Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97.
- Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38(3), 289–306.
- Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(s1), 285–310.
- 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.

- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.
- O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, 64(4), 715-748.
- R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- 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(3), 459-489.
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(4), 905-916.
- Zetterholm, E., & Abelin, Å. (2017). Swedish and Somali listeners' attitudes towards L2 Swedish speech. *Scandinavian Philology*, 15(2), 193-203.