Differences in second language speech fluency ratings: native versus nonnative listeners

Ralph L. Rose <rose@waseda.jp>
Waseda University Faculty of Science and Engineering

The perception of speech in cross-linguistic situations is known to be modulated by the listener's language background. In studies of the perception of foreign-accented speech by nonnative speakers (persons speaking in their nonnative language), nonnative listeners (persons listening to speech in their nonnative language) show greater comprehension than native listeners (persons listening to speech in their native language) (Bent and Bradlow 2003). Ratings of accentedness may also differ: Nonnative listeners judge foreign-accented speech less harshly than native listeners (Wester and Mayo 2014). Somewhat less studied is perceptions of fluency in similar cross-linguistic situations. Rossiter (2009) observed that fluency ratings by nonnative listeners of nonnative speech were lower overall than those given by native listeners. However, these rater groups were similar in that their ratings were most highly correlated with the speakers' articulation rate and silent pause frequency, suggesting that raters attend to these temporal parameters when rating fluency. In contrast, Foote and Trofimovich (2016) observed that native listeners depended more on silent pause frequency (over other factors) as a measure of fluency than did nonnative listeners, while the latter depended more on speech rate.

The diversity of language backgrounds in previous work could diminish effects associated with particular languages (cf., Bent and Bradlow's intelligibility benefit was larger for native-language sharing participants). Therefore, the present study narrows the scope of nonnative listener to those listeners who also share the same native language as the speaker. This may allow effects pertaining to one specific language group to be revealed. While narrower in scope, the present study is an attempt to contribute to the wider body of evidence on native vs. nonnative fluency ratings.

The present work takes advantage of a crosslinguistic speech corpus (Rose 2013) in which native speakers of Japanese spoke for several minutes in response to various elicitation tasks in both their native language and English, their second language. These tasks included reading aloud and two types of spontaneous speech tasks: picture description—in which participants described a single frame scene or multi-frame sequence of scenes— and topic narrative—in which participants were given a topic (e.g., explain basketball to someone who has never seen it)..

For each speaker in the corpus (N=35 adults), seven thirty-second clips of their English (i.e., nonnative) speech were extracted from the three task types (three clips each from picture description and topic narrative and one clip from reading aloud). These clips were rated on a 9-point scale for fluency (1=low, 9=high) by native listeners (N=34 native English-speaking adults) using the Amazon Mechanical Turk crowd-sourcing work system. Raters were instructed to judge the "smoothness" of the speech and not other features like pronunciation or syntactic complexity. These results were previously reported in a study of the relationship between first and second language speech and fluency ratings (Rose 2015). In the present study, this earlier work was extended by having nonnative listeners (N=20 native Japanese-speaking adults) listen to the same set of recordings and rate the fluency in a similar manner.

The present study also analyzes two additional temporal parameters, filled pauses and repairs. The central comparison is fluency ratings by nonnative listeners (who share the same native language as the speakers) to those by native listeners.

Results show that the nonnative listeners gave lower fluency ratings overall (mean=4.4, sd=1.6) than did native listeners (mean=4.9, sd=1.7). This difference was significant according to a repeated-measures anova [$F(1,52)=7.6$, $p<0.01$]. This is consistent with previous findings and supports the view that nonnative listeners tend to judge nonnative speakers' fluency more harshly, although they share a native language.

In order to examine which features of speech that raters are attending to in their judgments, a step-wise linear regression was performed with fluency rating as the dependent variable. Independent variables included articulation rate (syllables per minute of phonation time), silent pause length (mean length of silent gaps of 300ms or greater; c.f., De Jong and Bosker 2013), silent pause rate (number of silent gaps per minute), filled pause rate (number of occurrences of 'um' or 'uh' per minute), and repair rate (number of repair sequences per minute). The first three temporal parameters—articulation rate, silent pause length, and silent pause rate—were measured automatically using a Praat script (Boersma and Weenink 2013, Quené, Persoon, and De Jong 2011). The filled pause and repair rates were taken from transcriptions of the corpus. These variables plus elicitation task and rater language were included in the regression model.

The regression analysis shows that all factors were significant [$F(8,475)=118$, $p<0.001$; adjusted $R^2=0.66$]. Higher fluency rates are correlated with higher articulation rate, lower pause rate, shorter pause length, *higher* filled pause rate and lower repair rate. When separate analyses are performed on rater groups, articulation rate, silent pause length and filled pause rate remain for both groups while silent pause rate remains only for nonnative raters [$F(6,235)=104$, $p<0.001$; adjusted $R^2=0.72$] and repair rate remains only for native raters [$F(6,235)=58$, $p<0.001$; adjusted $R^2=0.58$].

The regression analysis further suggests differences between the reading aloud task and the other spontaneous speech tasks. After removing the reading aloud data, the overall models remain the same except that filled pause rate disappears from the optimal model for native raters [$F(3,203)=56$, $p<0.001$; adjusted $R^2=0.45$].

In contrast to previous findings, the present study suggests that nonnative raters judge fluency by somewhat different criteria than do native raters. While attention to articulation rate and pause length overlaps between them, nonnative raters also pay attention to pause rate while native raters pay attention to repair rate. This difference might be partially explained by differences in the perceptual tasks of the two raters. Previous research has shown that native listeners have a comprehension disadvantage (Bent and Bradlow 2003). Thus, more frequent pauses may actually aid their comprehension without being as intrusive as they are to nonnative listeners. On the other hand, once native listeners comprehend the speech, they are more likely to recognize repairs because it requires a higher level syntactic knowledge than some nonnative listeners have. Hence, the latter may simply be showing quicker word recognition, but lower syntactic comprehension than native raters.

The observed advantage for filled pause rate with nonnative raters may be related to

crosslinguistic differences in filled pause use. Filled pauses have been shown to be generally more frequent in Japanese than English (Watanabe and Toyama 2016). Hence, a higher filled pause rate may be regarded by these raters as somehow more natural, and therefore evidence of higher speaker fluency. But perhaps this becomes apparent to them only with the slower articulation rate of spontaneous speech.

Taken together with previous findings, the results here suggest that fluency ratings may be influenced by the native-language sharing status of listeners and raters. However, since the present study did not include non-native-language-sharing nonnative raters, a definitive conclusion cannot be made. This is the focus of future work.

References

Amazon Mechanical Turk web site. www.mturk.com

Bent, T. and Bradlow, AR. (2003) The interlanguage speech intelligibility benefit. In *Journal of the Acoustical Society of America*, 114(3): 1600-1610.

Boersma, P., Weenink, D. (2013): Praat: doing phonetics by computer [Computer program]. Version 5.4.03, retrieved 26 Dec 2014 from www.praat.org

De Jong, N. H., & Bosker H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In *Proceedings of The 6th Workshop on Disfluency in Spontaneous Speech (DiSS 2013)*, p. 17-20.

Foote, JA. and Trofimovich, P. (2016) A Multidimensional Scaling Study of Native and Non-Native Listeners' Perception of Second Language Speech. In Perceptual and Motor Skills, 122(2): 470-489.

Quené, H., Persoon, I., de Jong, N. Praat Script Syllable Nuclei v2 [Praat Script]. Version 28 Feb 2011, retrieved 26 Dec 2014 from https://sites.google.com/site/speechrate/Home/praatscript-syllable-nuclei-v2

Rose, R. (2013) Crosslinguistic Corpus of Hesitation Phenomena: A corpus for investigating first and second language speech performance. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, Lyon, France, pp. 992-996.

Rose, R. (2015) Temporal Variables in First and Second Language Speech and Perception of Fluency. In *The Scottish Consortium for ICPhS 2015 (Ed.), In Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

Rossiter, MJ. (2009) Perceptions of L2 Fluency by Native and Non-native Speakers of English. In *Canadian Modern Language Review*, 65(3):395.

Watanabe, M. and Toyama, S. (2016) Comparison of filled pause durations and frequencies in parallel corpus of English and Japanese. *Programme of The 16th Australasian International Conference on Speech Science and Technology (SST2016)*, p.42. Parramatta, Australia.

Wester, M. and Mayo, C. (2014) Accent rating by native and non-native listeners. In Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 7749-7753.