TEMPORAL VARIABLES IN FIRST AND SECOND LANGUAGE SPEECH AND PERCEPTION OF FLUENCY

Ralph ROSE

Waseda University
rose@waseda.jp

ABSTRACT

Evidence is accumulating that many temporal features of second language speech are correlated with those of first language speech. This study looks at the correlation between articulation rate, pause rate, and mean pause duration in Japanese first and English second language speech and how second language fluency raters perceive these. In a cross-linguistic corpus of spontaneous speech, mean pause duration was found to have a near-high correlation while the other two temporal variables have a moderate correlation. A subsequent elicitation of fluency judgments on the second language English speech via Amazon Mechanical Turk showed that ratings were highly dependent on pause duration, rather less on articulation rate, but not on pause rate. Results suggest that raters' perception of second language fluency is divergent from speakers' actual second language development: Ratings are related to features that are not indicative of second language development but rather of individual speech patterns.

Keywords: articulation rate, silent pause, fluency, second language acquisition

1. INTRODUCTION

A person who speaks relatively slowly in their native language is unlikely to suddenly speak much faster when they communicate in a second language. In this way, it is expected that many features of a person's articulation of their first and second languages will be similar. But this leads to a perceptual quandary: When an evaluator (e.g., a teacher or a testing coordinator) is judging the fluency of a second language speaker, which features of the speaker's speech should be used to judge how far the speaker has progressed toward specific fluency targets? Is it even possible to evaluate the fluency of a slow talker accurately? This paper reports on a research project that aims to understand how second language learners' fluency develops by focusing on the temporal features of their second language speech relative to those of their first language speech. After a review of the background literature, the paper reports on two experiments to address these questions—one a crosslinguistic corpus study and the other a fluency survey administered via Amazon Mechanical Turk—and then follows up with a discussion of the key results.

2. BACKGROUND

Much work has been done to look at the developmental variables of second language learners. Many researchers [1-7] have observed that as learners' proficiency in a language increases, their speech rate increases, their silent pause rate decreases, and the duration of their silent pauses decreases. Although these observations are reliable and have been replicated in other studies, many studies have not taken first language speech behaviour into account. For longitudinal studies that look at specific changes over time, this is perhaps not a problem. However, many of these studies are isochronic studies, taking a sample of many learners and assuming that scalar differences across the learners correlate with developmental differences over time. Thus, it is difficult to conclude whether these developmental trends in the temporal variables of second language speakers' speech is due to actual developmental changes or rather to individual differences in speech patterns.

A consensus is developing that in order to accurately evaluate second language speech patterns, it is useful to look at first language speech patterns in order to provide a baseline for interpretation. Several studies [8-10] have shown that some temporal features of speech are highly correlated between first and second language speech production. In particular, De Jong et al [11] have shown that articulation rate (which they operationalize as syllable duration—i.e., ratio of phonation time to number of syllables) in second language alone is not as good a predictor of second language fluency as when the measure is corrected by observations of first language articulation rate.

Their study focused on the relationship between (1) learners' cognitive ability to process language in a timely manner and (2) the overt temporal features of their speech production as evidenced by such things as articulation rate, silent pauses and filled pauses (e.g., *uh*/*um*). Segalowitz [12] referred to these two concepts respectively as cognitive fluency

and utterance fluency, which can be understood as properties of the speaker. Separate from these, Segalowitz also defined perceptual fluency as a property of the listener and how they perceive the speaker's speech to be fluent or otherwise. Utterance fluency is determined in part by a second language speaker's cognitive fluency in the language but also in part by individual speech patterns that are more clearly evidenced in their first language. But it is not clear how and to what degree listeners distinguish this when judging the fluency of a speaker. The goal of the present study is first, to look at how temporal features correlate between first and second language in order to understand the utterance fluency of speakers' second language speech and then second, to see how measures of perceptual fluency relate to these various temporal features. The next section describes two experiments to address these respective goals.

3. EXPERIMENTS

The two experiments described below both make use of a crosslinguistic corpus of spontaneous speech which consists of recorded elicitations of speech from individual speakers in both their first language (Japanese) and their second language (English) using parallel tasks in each language: reading aloud, picture description, and topic narrative. The corpus consists of over 11 hours of speech and over 60,000 words from 35 different participants.

For the present study, three temporal variables were selected for investigation: articulation rate, silent pause rate, and mean silent pause duration. Some comment on each of these is warranted. Articulation rate is defined here as the total number of spoken syllables divided by the total phonation time. In the past, some studies have studied speech rate (that is, words or syllables per total time), but that is arguably a composite measure of temporal variables including both articulation rate as well as silent pause rate and duration. Researchers recently use articulation rate as a non-confounding temporal variable.

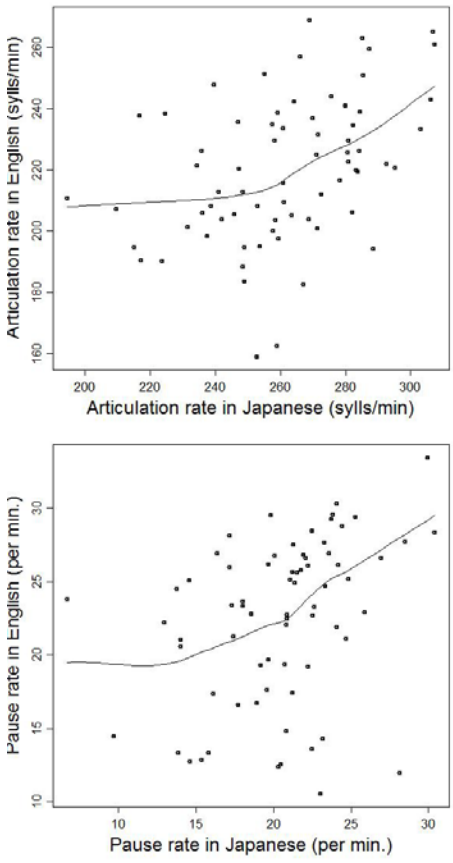
Silent pauses in this study consist of pauses that are longer than 300ms. The cut-off point for silent pauses has varied over the decades of pausological research [13], but in recent years, researchers have settled on a cut-off around 250-300ms. In this study, then, silent pause rate is the number of silent pauses per minute and the mean pause duration is the total silent pause time divided by the number of pauses. Measurements of these temporal variables were extracted using a Praat [14] script by Quené, Persoon, and de Jong [15].

2.1. Experiment 1: Temporal variables in first and second language speech

The purpose of the first experiment is to investigate the relationship between first and second language speech production in the corpus. While it is theoretically possible to talk about fluency in both read speech and spontaneous speech, some studies [e.g., 1] have shown that there are significant differences between the two at the utterance fluency level. Furthermore, spontaneous speech is the mode in which many second language learners are often tested and in which they strive to gain fluency. Thus, in this study, only the spontaneous speech recordings (i.e., picture description and topic narrative) were used.

Two recordings in both Japanese (first language) and English (second language) were analysed from each of the 35 speakers in the corpus. Because the mean pause duration was asymptotically distributed, the log value was used. The data comparing first and second language are shown in Figure 1 along with lowess curves and the Pearson product-moment correlations are shown in Table 1.

Figure 1: Articulation rate, silent pause rate, and (log) mean pause duration in Japanese (first language) and English (second language).



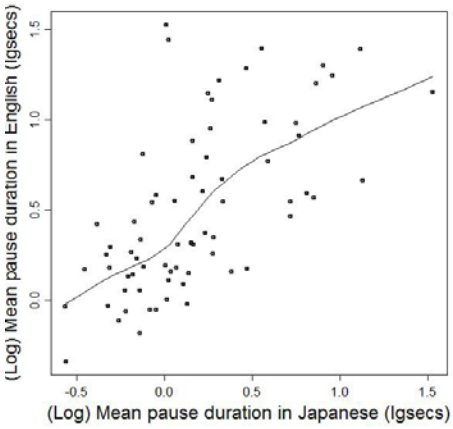


Table 1: Pearson product-moment correlations between English and Japanese temporal variables.

	r	p	r^2
Articulation rate	0.413	<0.001	0.17
Silent pause rate	0.341	<0.005	0.12
(Log) Mean silent pause duration	0.636	<0.001	0.40

Results show that although all three correlations are significant, silent pause duration has the highest correlation between first and second language and it is the only correlation that approaches a strong correlation (i.e., ≥0.7). Thereafter, articulation rate and then silent pause show moderate correlations.

2.2. Experiment 2: Perceived fluency of temporal variables in second language speech

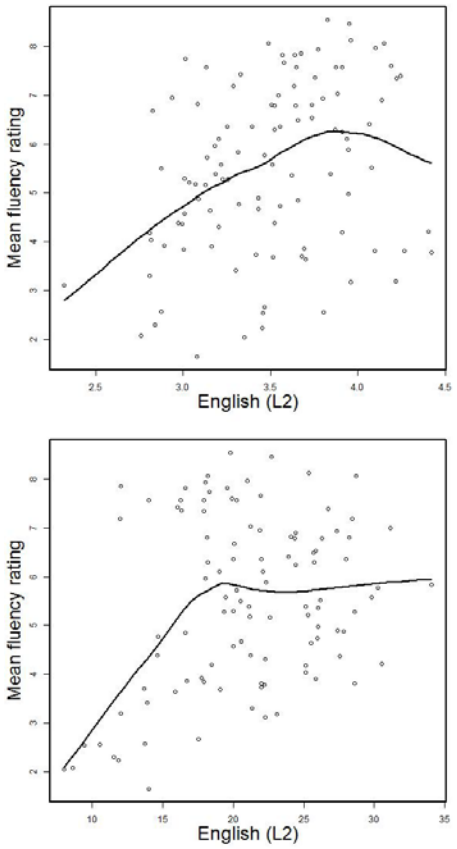
The purpose of the second experiment was to see how these three temporal variables in second language speech influence the perception of fluency by native speakers of the second language. For this purpose, 30-second samples of speech were extracted from the corpus in a pseudo-random manner (see [9] for comments on the sufficiency of 30-second samples for fluency evaluation). For each speaker, seven clips (three picture description, three topic narrative, and one reading aloud) were chosen. None of these clips were at the beginning or ending of the original recordings in order to avoid start-up or wrap-up effects. These clips were organized into seven balanced groups of 35 clips.

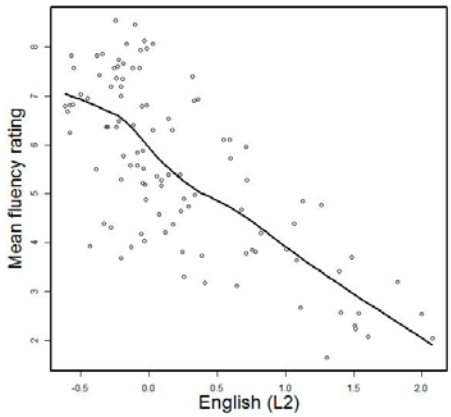
Because of the large size of the task, the data-gathering process took advantage of Amazon.com's Mechanical Turk [16] service—a system in which human intelligence tasks (HITs) are assigned to workers (so-called "Turkers") who do the tasks for the offered remuneration. More and more language-oriented investigations are taking advantage of this service for various research purposes [17].

For the present experiment, one HIT consisted of listening to one group of 35 clips and rating the fluency of each speaker on a 9-point Likert scale. Workers were given some examples to listen to, but were not instructed about how to judge fluency except that they should focus on "smoothness", and not on things like grammatical accuracy or clear pronunciation. In order to ensure that workers did the task conscientiously, a background script kept track of the play and pause time-stamps of each audio component. In addition, some "attention check" items were added in which they were instructed to select a specified rating. Finally, the reading aloud clips were included to check rating consistency since it was predicted that these clips would be typically perceived as more fluent than spontaneous speech clips. The fluency evaluations for the spontaneous speech clips are not included in the following analysis, however. Finally, the order of presentation of the 35 clips was randomized in every HIT (thus, a different order for each worker).

108 workers completed 350 HITs (50 HITs in each of 7 groups). Many workers completed more than one group and 34 workers completed all seven groups. There is no difference in the results below if the analysis is limited to just the data from these 34 workers. Therefore, the analysis below includes data from all workers. Cumulative results for the seven groups are shown in Figure 2 with lowess curves.

Figure 2: Relationship between articulation rate, silent pause rate, and (log) mean pause duration in second language (English) speech and fluency ratings (1=low to 9= high).





The three temporal variables were taken as factors in a linear model with fluency rating as the dependent variable. The optimal results of a stepwise regression on this model are shown in Table 2.

Table 2: Stepwise linear regression with temporal variables as predictive factors and fluency rating as dependent variable.

	Est$_\circ$	Std. Error	t	p
(Intercept)	2.4981	1.0021	2.493	<0.05
Articulation rate	0.9759	0.2854	3.419	=0.001
(Log) mean silent pause duration	-0.5950	0.0820	-7.257	<0.001
Adjusted R^2 = 0.4711; F(2,67) = 31.73, p<0.001				

The two variables that showed the highest correlation between first and second language speech—mean silent pause duration and articulation rate—end up in the optimal model, in the same order of importance as their correlations in Experiment 1. In contrast, the factor with the lowest correlation between first and second language—pause rate—has disappeared. In short, these results show that second language fluency evaluators' ratings are particularly sensitive to the temporal variables which are closely linked to the speakers' first language speech performance.

4. DISCUSSION

The results of the two experiments are somewhat surprising in that they suggest that fluency evaluators are judging fluency based on the wrong things: That is, it would seem that their ratings tell us more about each individual's speaking style (regardless of language) rather than the current state of their second language performance level. This has two practical implications. First, if a second language learner's aim is to be perceived as a fluent speaker of the language, they need to focus primarily on minimizing pause duration and secondarily on articulating faster, even if this is inconsistent with their normal speech style. While challenging, this may be feasible for some limited speech modes such as rehearsed oral presentations. Future work could focus on confirming this hypothesis: whether such pseudo-spontaneous speech is actually evaluated as more fluent than authentic spontaneous speech.

Second, where it is important that fluency ratings actually reflect second language development (e.g., in second language skill assessment), then it is crucial that raters understand which features of a speaker's speech are actually indicative of their development. In the case of spoken fluency, the results here suggest that silent pause rate is more aligned with second language development than articulation rate and silent pause duration.

The regression model described in the results above explains about 47% of the observed variance meaning that a number of other factors not looked at (yet) in this study are also at play in fluency judgments. These might include filled pauses (e.g., in English, *uh* and *um*), repeated words, self-corrections and clipped words, as well as lengthenings (words with one or more elongated segments). While analysing these is relegated to future work, lengthenings warrant some further comment here. To some extent, the results may provide some evidence that lengthenings influence fluency judgments in that articulation rate, which is confounded with lengthenings, does explain some of the observed variance: It could be the case that fluency raters are being influenced specifically by the lengthenings rather than by the individual speaker's normal articulation rate. Further analysis will be necessary to determine if this is the case.

5. CONCLUSION

This study has added to the growing body of evidence that many temporal features of second language speech performance are correlated with those of first language speech behaviour and thus may not reliably indicate second language fluency development on their own.

6. REFERENCES

[1] Cucchiarini, C., van Doremalen, J., Strik, H. (2010). Fluency in non-native read and spontaneous speech. *Proc. of Disfluency in Spontaneous Speech (DiSS) and Linguistic Patterns in Spontaneous Speech (LPSS) Joint Workshop* (Tokyo, Japan), 15-18.

[2] Kormos, J. Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145-164.

[3] Riazantseva, A. (2001). Second language proficiency and pausing a study of Russian speakers of English. *Studies in Second Language Acquisition*, 23(4), 497-526.
[4] Rieger, C. L. (2003). Disfluencies and hesitation strategies in oral L2 tests. *Proc. of DiSS '03, Disfluency in Spontaneous Speech Workshop* (Göteborg, Sweden), 41-44.
[5] Tavakoli, P. (2011). Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*, 65(1), 71-79.
[6] 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, 1-30.
[7] Trofimovich, P., Baker, W. (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, 28(2), 251-276.
[8] Cox, T., Baker-Smemoe, W. (2012). The relationship between L1 fluency and L2 fluency across different proficiency levels and L1s. Presentation at Workshop Fluent Speech (Utrecht University, The Netherlands).
[9] Derwing, T. M., Munro, M. J., Thomson, R. I., Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(4), 533-557.
[10] De Jong, N., Steinel, M. P., Florijn, A., Schoonen, R., Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34(5), 893-916.
[11] De Jong, N., Groenhout, R., Schoonen, R., Hulstijn, J.H. (in press). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behaviour. *Applied Psycholinguistics*.
[12] Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
[13] De Jong, N., Bosker, H.R. (2013) Choosing a threshold for silent pauses to measure second language fluency. *Proceedings of DiSS '13, Disfluency in Spontaneous Speech Workshop* (Stockholm, Sweden), 17-20.
[14] Boersma, P., Weenink, D. (2013): Praat: doing phonetics by computer [Computer program]. Version 5.4.03, retrieved 26 Dec 2014 from www.praat.org.
[15] 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/praat-script-syllable-nuclei-v2.
[16] Amazon Mechanical Turk web site. www.mturk.com
[17] Schnoebelen, T., Kuperman, V. (2010) Using Amazon Mechanical Turk for linguistic research. *Psihologija*, 43(4), 441-464.