Preceding Word Information for Predicting Speech Errors in EFL Speech

Ki Woong Moon
The University of Arizona (USA)

Speech errors, including disfluency errors (e.g. filled pauses (“uh”, “um”), repetition (“I me-mean right now.”)), and mispronunciation of speech segments (e.g. “think” as /sɪŋk/) are natural occurrences in speech production, especially for second language (L2) speakers. The presence of these errors affects speech comprehensibility [1] and consequently affects speech fluency and proficiency [2]. From the perspective of L2 learning, detecting these errors is important, but the process requires human intervention [3,4] which has drawbacks related to cost and time.

Finding a feature that efficiently predicts word-level speech errors would be important to reduce the cost and time of examining speech errors embedded in L2 speech. This study investigates the usefulness of preceding word information to help predict speech errors in non-native speakers’ read speech. Non-native speakers experience an additional cognitive load when working in their second language (L2) [5], possibly due to the less automatic nature of grammatical encoding in an acquired language [6]. Based on the greater cognitive load of non-native speech production, the study hypothesized that L2 speakers may face an additional working memory load because they need to process two tasks at once--what to produce now, and what will be produced next. The current study utilized two features of the preceding word to determine whether they help to predict errors in the current word: mean syllable duration of the preceding word and the number of syllables of the preceding word. The durational feature may help capture the additional cognitive load experienced by L2 speakers in producing speech, as increased cognitive load can lead to more frequent and longer silent or filled pauses [6], raising word duration.

The study conducted descriptive and inferential statistical analyses of two L2 read speech corpora: Prawn_dB [7] (native speakers of Korean) and L2-ARCTIC [8] (native speakers of Arabic, Hindi, Korean, Mandarin, Spanish, and Vietnamese), that have different types of errors annotated (disfluency, mispronunciation). In the study, all kinds of words were considered regardless of their categories. Results (Table I) suggest that in addition to the significant effects of present word information when the preceding word has a longer average syllable duration, there is an increased likelihood of a mispronunciation error occurring in the current word. Figure 1 shows the relationship between the predicted probabilities of present and preceding word’s mean syllable duration. In addition, the results of the model comparison showed that adding the preceding word information produced a significantly better fit of the mispronunciation data ($\chi 2\ (1) = 7.89$, $p<.005$). The study also conducted statistical analyses for word frequency and word category effects, but no significant effects were found in word frequency and word categories.

The study also investigated the characteristics of the target word itself (current word) on whether that word contains an error. The results in both corpora showed that the mean syllable duration of the present word is likely to be longer if a speech error exists in the word. Also, the study found that words with a higher number of syllables tend to contain speech errors in them.

In summary, the results suggest that the average duration of the preceding word has a significant effect on predicting mispronunciation errors in L2 read speech. In addition, both syllable duration and the number of syllables of the current word correlate with the presence of errors. The correlation with current word syllable duration likely reflects speech slowing at errors (since filled pauses and pauses themselves add to the average syllable duration). The effect of the number of syllables of the current word likely reflects L2 speakers’ difficulty processing polysyllabic (and potentially low-frequency) words. These findings can facilitate the future construction of a system for automatically detecting L2 speech errors by incorporating the preceding word’s information as one of the linguistic features.

Keywords: L2 speech; disfluency error; mispronunciation; preceding word information; speech error detection

Table 1. Summary of the final generalized mixed-effects logistic model. Fixed factors: the number of syllables in the present word (reference level: present word is a disyllabic word), the mean syllable duration of the present and preceding word for a categorical dependent variable of mispronunciation error (vs. normal word).

Fixed Factor	Estimate	SE	*z* value	*Pr*(>|z|)
(Intercept)	-0.857	0.149	-5.762	<0.001
Mean syllable duration (present word)	1.573	0.233	6.738	<0.001
Mean syllable duration (preceding word)	0.522	0.184	2.830	<0.005
Present word is a monosyllabic word ('1')	-1.192	0.094	-12.607	<0.001
Present word has three or more syllables ('3+')	0.704	0.117	6.036	<0.001

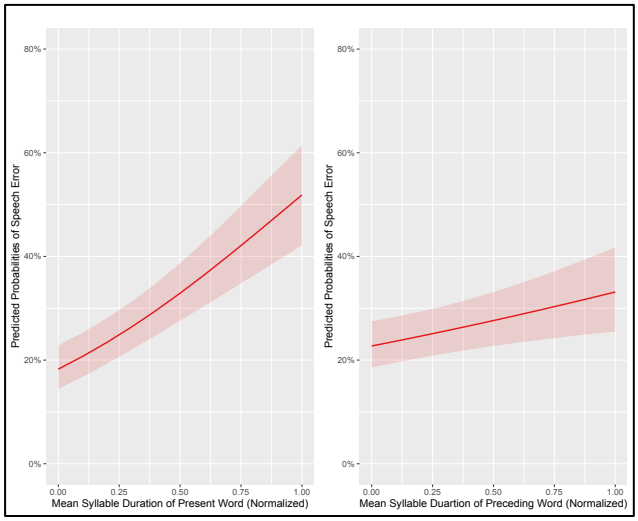

Fig 1. Predicted probabilities of speech errors for continuous fixed factors. The predicted probabilities in both graphs increase as the mean syllable duration of the present and preceding word becomes higher. The steeper slope and lower variability in the predicted probabilities associated with the mean syllable duration of the present word (left) suggest that it has a stronger predictive power compared to the mean syllable duration of the preceding word (right).

References

[1] Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15, 905–916.

[2] Park, S. (2021). Human and machine judgment of non-native speakers' speech proficiency [Doctoral dissertation, University of Arizona].

[3] Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach, *Language Learning*, 40(3), 387-418.

[4] Moon, K. W. (2022). Automatic detection of word-level suprasegmental speech errors in EFL utterance [Master's thesis, Hankuk University of Foreign Studies]. http://hufs.dcollection.net/common/orgView/200000628701

[5] Fehringer, C. & Fry, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory. Folia Linguistica,41(1-2), 37–72.

[6] Christodoulides, G. (2016). Effects of cognitive load on speech production and perception [Doctoral dissertation, UCL-Université Catholique de Louvain]. https://dial.uclouvain.be/pr/boreal/object/boreal:177060/datastream/PDF_01/view

[7] Chung, H., Jang, T. Y., Yun, W., Yun, I. & Sa. J. (2008). A study on automatic measurement of pronunciation accuracy of English speech produced by Korean learners of English. *Language and Linguistics*, 42, 165-196.

[8] Zhao, G., Sonsaat, S., Silpachai, A., Lucic, I., Chukharev-Hudilainen, E., Levis, J., Gutierrez-Osuna, R. (2018) L2-ARCTIC: A non-native English speech corpus. *Proceedings of the Interspeech 2018*, (pp.2783-2787), doi: 10.21437/Interspeech.2018-1110