

Disfluency behavior over time in non-contemporaneous L2 English voice messages

Patrik Hrabánek, *ACLCL, University of Amsterdam, The Netherlands*

Willemijn Heeren, *LUCL, Leiden University, The Netherlands*

Human speech inherently exhibits disfluencies, as evidenced by the presence of e.g. filled and unfilled pauses, interruptions, and repetitions. Previous research has shown that disfluency behavior can be speaker-specific, e.g. [1-3]. Because of this, between-speaker and within-speaker variation in fluency are very relevant in forensic speaker comparisons (FSCs), where the analysis of disfluency behavior can be used to help assess whether the speaker on an incriminating recording and the suspect may be the same speaker.

As forensically relevant research into disfluencies in crosslinguistic environments is still rather scarce (although see [4, 5]), the present study investigated disfluency distributions in L2 English speech. Existing studies on disfluency behavior showed that speakers tend to differ in their disfluency behavior based on speech task (i.e. conversation, voicemail, monologue) [2], however, possible changes in disfluency behavior within the same speech task over time were not investigated. This led us to ask the following research question: Is the production of disfluencies consistent across time within individual speakers?

To answer this question, we examined speech in voice messages transmitted through WhatsApp, which was chosen due to the ever-increasing use of messaging apps over traditional voicemail. The analysis was carried out using the TOFFA framework [1], a forensic disfluency analysis tool used in real forensic casework, which we slightly modified based on observations from our data; we added a category of filled pauses (FPs) surrounded by unfilled pauses on both sides (see Table 1). Crucially, we used non-contemporaneous recordings (i.e. recordings made on two different occasions), as real casework typically requires examining multiple recordings captured at various points in time.

The participants were 26 female L1 Dutch speakers, who were instructed to record two different minute-long voice messages in their L2 English set apart by at least one week. They recorded themselves speaking about one of five possible topics (e. g. favorite hobbies, personal travelling experiences etc.) in a quiet room holding their phone directly in front of their mouth, in order to simulate a real-life scenario. Seven days after the first message, a reminder was sent to facilitate the recording of the second message. Both recordings were manually annotated for disfluencies in Praat [6] by two coders, each processing either the first or second message of the first half of the participants, and switching the message annotated for the other half. The statistical analysis was done using linear mixed-effects models built in R using the *lme4* package [7]. Disfluency counts were expressed as number of occurrences per minute since individual messages differed in their durations (mean = 65.4 seconds; SD = 11.5 seconds).

Speakers on average produced 42.7 disfluencies per minute (see Figure 1). There was a small statistically significant effect of Message (first or second), explained only by the category of prolongations. Counts of other disfluency categories did not statistically differ between messages. Between-speaker variation in the distribution of FPs was found, signaling that some speakers very often surround their FPs with unfilled pauses on both sides, whereas others consistently link their FPs to the rest of their utterance. Speakers also tended to use unfilled pauses more frequently before and after FPs that were followed by nasals, rather than those that consisted only of vowels. Acoustic analysis of the FPs showed no significant differences between messages in terms of F0 and the first three formants, which concurs with previous literature [5, 8].

All in all, our results point to stability in L2 disfluency behavior over shorter periods of time, and also show individual variation between speakers. This finding is valuable for forensic applications, as well as for measuring L2 fluency more generally.

Table 1: Modified TOFFA categories used in this study. Colors correspond to those in Figure 1.

Categories	Subcategories + examples
Unfilled pauses	unfilled pause (> 200ms) [po]
Filled pauses	FP [er]
	FP + nasal [erm]
	unfilled pause + FP + unfilled pause [er-po]
	unfilled pause + FP + nasal + unfilled pause [erm-po]
Prolongations	vowel, nasal, or lateral prolongation [prov]
	fricative or plosive prolongation [proc]
Repetitions	part-word repetition [pwr] <i>it's r-really far</i>
	whole-word repetition [wrep] <i>and he- he was a policeman</i>
	phrase repetition [prep] <i>turn right by the- by the department store</i>
Interruptions	word interruption [wint] <i>he took me to lun- dinner</i>
	phrase interruption [pint] <i>I was going to- I left at five</i>

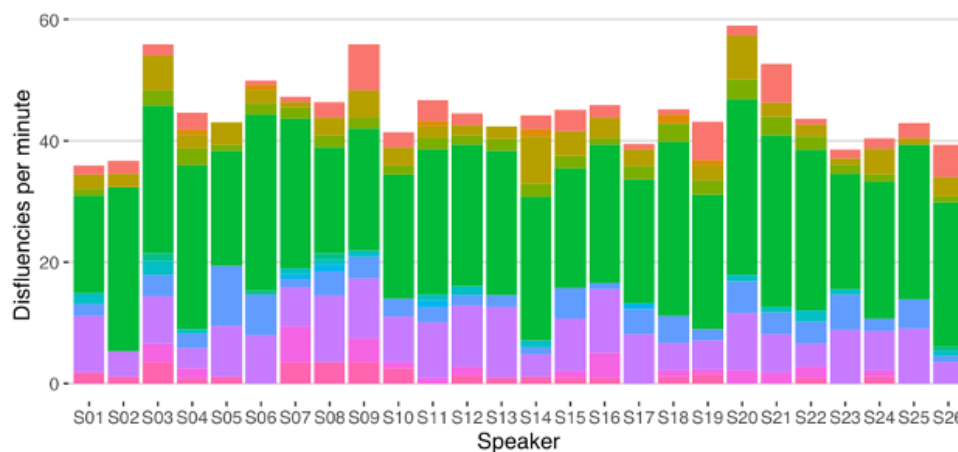


Figure 1: Relative frequency of all disfluency types per minute for each speaker (both messages combined). The color legend can be found in Table 1.

References:

- [1] McDougall, K. & Duckworth, M. (2017). Profiling fluency: An analysis of individual variation in disfluencies in adult males. *Speech Communication*, 95, 16-27.
- [2] Harrington, L., Rhodes, R., & Hughes, V. (2021). Style variability in disfluency analysis for forensic speaker comparison. *The International Journal of Speech, Language and the Law*, 28(1), 31–58.
- [3] Braun, A., Elsässer, N. & Willems, L. (2023). Disfluencies revisited—Are they speaker-specific? *Languages*, 8, 155-185.
- [4] de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36(2), 223–243.
- [5] de Boer, M. M., Quené, H. & Heeren, W. F. L. (2022). Long-term within-speaker consistency of filled pauses in native and non-native speech. *JASA Express Letters*, 2(3), 035201.
- [6] Boersma, P. & Weenink, D. (2023). PRAAT: doing phonetics by computer. Computer program. version 6.4. Available online at: <http://www.praat.org/>.
- [7] Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using ‘lme4’. *Journal of Statistical Software*, 67(1), 1–48.
- [8] Hughes, V., Wood, S., & Foulkes, P. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law*, 23(1), 99–132.