EasyChair Preprint

№ 3842

Contribution of Disfluencies to Perception of Speech Quality

Katie Brewer, Ai Leen Choo and Sarah Smith

July 12, 2020

Contribution of Disfluencies to Perception of Speech Quality

Katie Brewer[1], Ai Leen Choo[1], and Sara Smith[2]

[1]Department of Communication Sciences and Disorders, Georgia State University

[2]Department of Teaching and Learning, University of South Florida

Author Note

The authors declare that there are no conflicts of interest with respect to this preprint.

Correspondence regarding this paper should be addressed to Katie Brewer. Email: kbrewer15@student.gsu.edu

Abstract

Purpose: To examine the effect of disfluencies on perceptions of speech quality.

Method: Audio samples were gathered from 20 bilingual speakers and rated by unsophisticated listeners using subjective fluency and quality measures.

Results: Ratings were highest for speakers with a higher number of disfluencies. Analyses indicate a significant difference between the low and average rated groups.

Conclusions: Higher disfluencies may contribute to better listener recall and positive listener perceptions of speaker effort for bilingual speakers.

Keywords: disfluencies, bilinguals, ratings, perception

Contribution of Disfluencies to Perception of Speech Quality

Fluency and disfluency exist on a continuum of speech production (Starkweather & Givens-Ackerman, 1997). Fluency is commonly defined as the absence of disruptions in the forward flow of speech. Disfluencies, or speech interruptions, include interjections (e.g., uh, um), hesitations, pauses, repetitions, blocks, and prolongations (Ambrose & Yairi, 1999). In the field of communication sciences and disorders (CSD), disfluencies are classified into two types: (1) stuttering-like disfluencies (SLD) and (2) typical disfluencies (TD). A high rate of SLDs, which include part-word repetitions, blocks, and prolongations, is indicative of a fluency disorder (Ambrose & Yairi, 1999). Whereas interjections, pauses, and hesitations are considered more normal disfluencies and are common in the speech of normally fluent individuals (Ambrose & Yairi, 1999).

Findings related to perceptions of disfluent speakers have been mixed, although most studies indicate that listener perceptions of highly disfluent speakers tend to be overwhelmingly negative. Susca and colleagues (Panico, Healey, Brouwer, & Susca, 2005; Susca & Healey, 2002) reported that more disfluent speakers are generally viewed more negatively than normally fluent speakers. Individuals with high levels of disfluencies in their conversational speech are perceived as less competent (i.e., poor grasp of the content or subject) than fluent speakers. Highly disfluent speakers may also be viewed as less intelligent compared to normally fluent speakers (Susca & Healey, 2002). Speech with high levels of disfluency has been reported to induce higher levels of discomfort in the listener compared to fluent speech. This discomfort may correlate with the listener's need to employ active listening in order to comprehend a disfluent speaker (Susca & Healey, 2002). Listeners have been observed to react to this discomfort with impatience and frustration at their inability to follow the narrative

(Panico, Healey, Brouwer, & Susca, 2005). In addition, listeners judged highly disfluent speech as more unnatural compared to speech with fewer disfluencies. This effect was especially pronounced when listeners were able to see the speaker and moments of disfluency compared to audio-only samples where the speaker is not visible (Martin & Haroldson, 1992).

Several studies found that listeners exhibited poorer information recall when the speech samples contained disfluencies (Cyprus, Hezel, Rossi, & Adams, 1984; Hulit, 1976; Panico & Healey, 2009). Listener recall was especially affected when disfluencies occurred on high information words (i.e., key words). While Cyprus, Hezel, Rossi, & Adams (1984) found that mild stuttering (i.e., speech samples with 5% disfluencies) did not affect listener recall or comprehension, Panico & Healey (2009) reported that disfluencies at any severity level, but most notably part-word repetitions, whole-word repetitions, and prolongations, diminished listener recall and comprehension. Conversely, Fraundorf and Watson (2011) found that listeners showed better recall when speakers inserted interjections (i.e., um, uh) during storytelling. It is important to note that the Fraundorf and Watson (2011) study did not include SLDs or speakers with fluency disorders, in contrast to previous studies (e.g., Cyprus, Hezel, Rossi, & Adams, 1984; Panico & Healey, 2009). Finally, Pytko and Reese (2013) reported that interjections (i.e., um, uh) in the speech of normally fluent speakers affected perceptions of speakers' preparedness and ease of comprehension rather than perceptions of the speaker's intelligence.

The ambiguity surrounding the relationship between disfluencies and perceptions of speakers point to the need for more research. The aim of this study was to examine the effect of disfluencies on perceptions of speech quality. Based on previous findings, we expected speech

samples with a higher number of disfluencies to be rated lower on speech quality than samples with fewer disfluencies.

Method

Participants

Participants in the study were speakers who provided audio-only speech samples and listeners who rated the fluency and quality of the audio samples. The speakers were 20 adult Spanish-English heritage non-balanced domain bilinguals (14 female speakers and six male speakers) with a mean age of 28.19 years (*SD* = 7.85). Based on self-report measures, all speakers learned Spanish as their first language (L1) from birth and learned English in school (L2) between ages 3-7 years. All speakers were normally fluent (i.e., did not have a fluency disorder) and spoke English as their dominant language in academic and professional experiences. The raters were monolingual and bilingual adults who spoke English as their dominant language.

Speakers. Audio samples were collected using a simulated video conference "virtual meeting" format, in which each speaker responded to six pre-recorded workplace-related questions presented by multiple interviewers. Examples of the questions include: "Tell me about a time you had too many things to do and had to prioritize. How did you organize your time?" Speakers were given 2 minutes and 30 seconds to respond to each interview question.

Raters. A total of 222 raters were asked to rate the audio samples based on subjective language fluency and quality measures on a scale of 1 (low quality) to 5 (high quality). After listening to the speakers' responses, the raters were asked to answer questions regarding their perceptions of the speaker (e.g., "How positively did you view the [speaker]?") Each response was rated by at

least two raters. The rating scores for each of the six responses were averaged to produce a single score for each speaker.

Transcription and coding of disfluencies

Audio samples were transcribed and coded using Computerized Language Analysis (CLAN), a program under the Child Language Data Exchange System (CHILDES; MacWhinney, 2003) in three passes. For the first pass, the samples were transcribed by two researchers. In the second pass, disfluencies were identified and coded using CLAN by a researcher (KB) with extensive experience identifying disfluencies. The types of disfluencies coded were prolongations, broken words, blocks, repeated segments (i.e., part-word repetition), phonological fragments, whole-word repetitions, phrase repetitions, phrase revisions, pauses, and filled pauses (i.e., interjections; Ratner & Brundage, 2020). The total number of disfluencies was calculated using FLUCALC, a program under CLAN. For the third pass, 10% data (i.e., 12 speech samples) were randomly selected to be recoded and analyzed by a fourth researcher who was trained to identify disfluencies. Interrater reliability for disfluency coding was high (ICC = 0.93).

Ratings group

Ratings were grouped into three categories: low, average or high. Ratings which were one SD (0.577) or more below the mean (M = 3.362) were coded into the low ratings group, ratings which were one or more SD above the mean were coded into the high ratings group, and ratings between these two scores were coded into the average rating group.

Data analysis

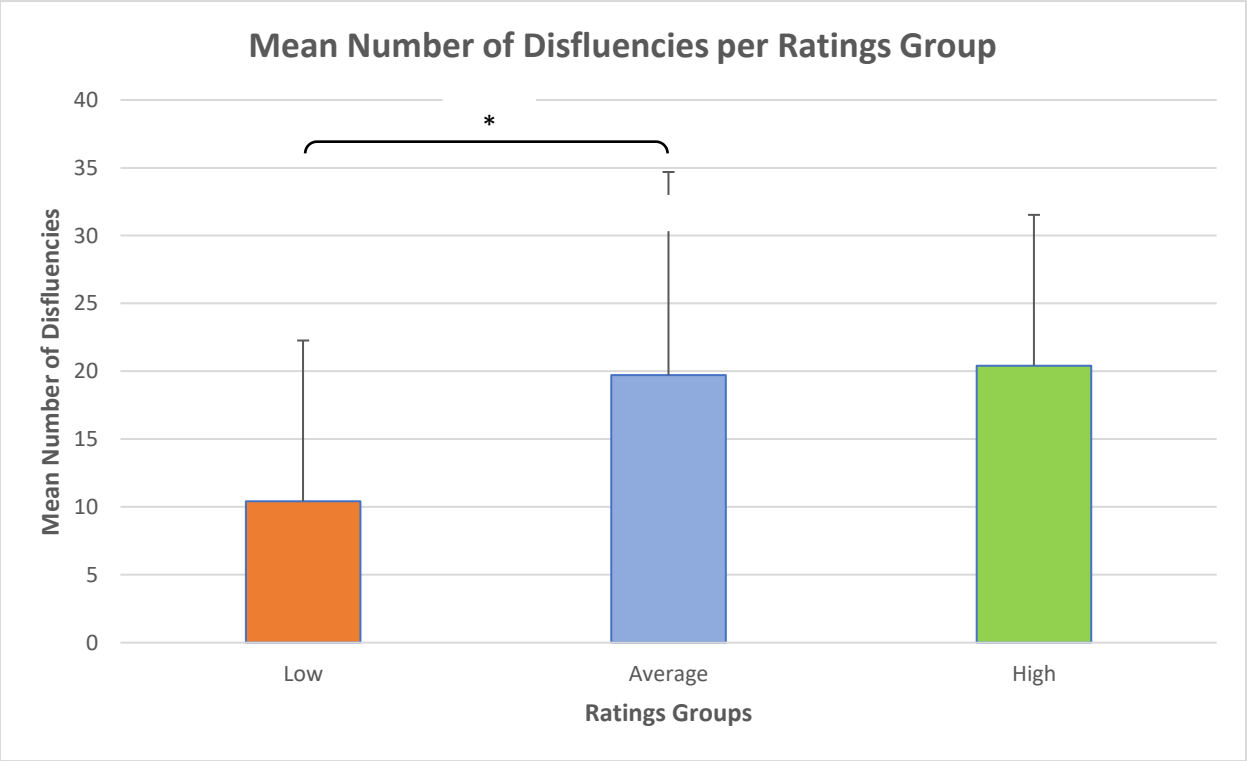
A nonparametric test was conducted to determine the relationship between the total number of disfluencies produced and ratings received (*low*, *average*, *high*).

Results

Results showed a difference in the number of disfluencies and ratings received. A Kruskal-Wallis H test showed that the assigned ratings group was significantly correlated with the total number of disfluencies produced by the speaker ($H[2] = 13.554$, $p = .001$). Speakers who received high ratings had more disfluencies ($M = 20.41$, $SD = 11.12$) in their speech compared to speakers who received average ($M = 19.72$, $SD = 14.97$) or low ratings ($M = 10.42$, $SD = 11.85$). Post-hoc analyses (Bonferroni corrected) indicate a significant difference between the *low* and *average* rated groups ($p < 0.05$). No other comparisons were significant. Further analysis revealed that the most common type of disfluency was filled pauses, or interjections ($M = 4.44$, $SD = 2.92$).

Figure 1

Mean Number of Disfluencies per Ratings Group.



Discussion

Current findings did not support our hypothesis that speech samples with a higher number of disfluencies would receive lower ratings than samples with fewer disfluencies. Our findings also did not support previous research in which listeners exhibited negative perceptions towards speakers with more disfluencies (e.g., Panico, Healey, Brouwer, & Susca, 2005). Several factors may account for this discrepancy, including the types and frequency of disfluencies present in the audio samples.

It is crucial to note that the disfluencies coded in the present study included both SLDs and TDs. This differs from other studies that measured SLDs separately from TDs (e.g., Ambrose & Yairi, 1999) or only examined one disfluency category (e.g., Susca & Healey, 2002). In our study, the most common type of disfluency in the speech samples was filled pauses or interjections, which are considered TDs. This type of disfluency is similar to those measured by Fraundorf and Watsons's (2011) study which found a correlation between recall and frequency of interjections. In contrast, studies that found negative listener perceptions (e.g., Panico & Healey, 2009) examined SLDs and individuals with fluency disorders. In addition to disfluency type, initial observations indicate that the majority of our speech samples contained less than 10% disfluencies. In the studies examining individuals with fluency disorders, researchers found that listener perceptions of the speaker were not negative until the levels of disfluencies reached or exceeded 10% (i.e., Panico, Healey, Bouwer, & Susca, 2005; Susca & Healey, 2002), after which listener perceptions became decidedly negative. These two factors, type and frequency of disfluencies, are likely the most meaningful factors that distinguish our findings from previous studies that reported a negative correlation between ratings and

disfluencies. The pattern was reversed in our data, as the highest number of disfluencies received the highest ratings.

Research related to improved recall and perceived effort may also explain our results. Past studies found that some types of disfluencies (i.e., interjections) increased listener recall (Fraundorf & Watson, 2011). In our study, it is plausible that raters experienced better recall in the higher disfluency samples and, accordingly, rated these speakers better (Barr, 2001; Foxtree, 2002).

Perception of effort may be another factor contributing to positive ratings for samples with high disfluency levels. Bilinguals typically experience greater communicative burden speaking in their L2 compared to their L1, and this burden may be reflected as higher levels of disfluencies (Dornic, 1980; Lim, Lincoln, Chan, Onslow, 2008). Higher disfluencies, particularly in bilinguals as in the current study, may be perceived as the speaker "trying harder" to speak their L2 and, as such, may be perceived more positively, receiving higher ratings. More research would be needed to confirm current findings.

There are several limitations in our study which should be taken into consideration for future studies. First, the speech samples were audio-only rather than audiovisual. Past studies comparing audiovisual and audio-only speech samples have noted a difference in perception between the two mediums. Specifically, audiovisual samples were consistently rated as more unnatural than audio-only samples (Martin & Haroldson, 1992). In addition, disfluency counts for audio-only samples tend to be lower than for audiovisual samples (Rousseau, Onslow, Packman, & Jones, 2008). Another limitation is that differences in ratings due to gender or sex were not considered during this study. Research indicates that perception judgements of others may vary based on gender or sex. Specifically, attitude measures have demonstrated that men are

more likely than women to utilize stereotypes and express racist or sexist attitudes (Nelson, Signorella, & Botti, 2016). The presence of more females than males in our study may have contributed to overall more positive ratings, though the raters' gender identification were not taken into account in the present study. Finally, the raters in this study were not given specific rating criteria for judging the speakers. The questions asked of the raters were broad and highly subjective (e.g., "How positively did you view the [speaker]?")

Future studies would benefit from a comparison of disfluency types (i.e., SLD and TD) and listener perception. A disfluency categorization scheme that measures the presence of SLDs and TDs would be appropriate for this purpose (e.g., Ambrose & Yairi, 1999). Future studies would also benefit from the inclusion of speakers with fluency disorders for comparison purposes.

References

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. Journal of Speech, Language, and Hearing Research, 42(4), 895-909.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. Oralité et gestualité: Interactions et comportements multimodaux dans la communication, 597-600.

Cyprus, S., Hezel, R. T., Rossi, D., & Adams, M. R. (1984). Effects of simulated stuttering on listener recall. Journal of Fluency Disorders, 9(3), 191-197.

Dornic, S. (1980). Language dominance, spare capacity, and perceived effort in bilinguals. Ergonomics, 23, 369–377.

Foxtree, J. E. (2002). Interpreting pauses and ums at turn exchanges. Discourse Processes, 34(1), 37-55.

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. Journal of Memory and Language, 65(2), 161-175.

Hulit, L. M. (1976). Effects of nonfluencies on comprehension. Perceptual and Motor Skills, 42(3_suppl), 1119-1122.

Lim, V. P., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Stuttering in English–Mandarin bilingual speakers: The influence of language dominance on stuttering severity. *Journal of Speech, Language, and Hearing Research, 51*(6), 1522-1537.

MacWhinney, B. (2003). Child language analyses (CLAN) (version 23 September 2003) [Computer software]. Pittsburgh, PA: Author.

Martin, R. R., & Haroldson, S. K. (1992). Stuttering and speech naturalness: Audio and audiovisual judgments. Journal of Speech, Language, and Hearing Research, 35(3), 521-528.

Nelson Jr, L. R., Signorella, M. L., & Botti, K. G. (2016). Accent, gender, and perceived competence. *Hispanic Journal of Behavioral Sciences*, *38*(2), 166-185.

Panico, J., & Healey, E. C. (2009). Influence of text type, topic familiarity, and stuttering frequency on listener recall, comprehension, and mental effort. Journal of Speech, Language, and Hearing Research.

Panico, J., Healey, E. C., Brouwer, K., & Susca, M. (2005). Listener perceptions of stuttering across two presentation modes: A quantitative and qualitative approach. Journal of Fluency Disorders, 30(1), 65-85.

Pytko, J. L., & Reese, L. O. (2013). The Effect of Using" Um" and" Uh" on the Perceived Intelligence of a Speaker. College of St. Elizabeth Journal of the Behavioral Sciences.

Ratner, N., & Brundage, S. (2020). In A clinician's complete guide to CLAN and PRAAT. Retrieved from https://talkbank.org/manuals/Clin-CLAN.pdf

Rousseau, I., Onslow, M., Packman, A., & Jones, M. (2008). Comparisons of audio and audiovisual measures of stuttering frequency and severity in preschool-age children. American Journal of Speech-Language Pathology.

Starkweather, C. W., & Givens-Ackerman, J. (1997). Stuttering: Pro Ed.

Susca, M., & Healey, E. C. (2002). Listener perceptions along a fluency–disfluency continuum: A phenomenological analysis. Journal of Fluency Disorders, 27(2), 135-161.