How to Um Get a Job: Disfluency Rates as Predictors of Interview Success

Amare N. Allen,[1,2] Menton E. Deweese,[1] Doriana Lacitignola,[2] Duane G. Watson[2]

[1]*The School for Science and Math, Vanderbilt University, Nashville, Tennessee, U.S.A 37235*
[2]*Department of Psychological Sciences, Vanderbilt University Nashville, Tennessee, U.S.A 37235*

KEYWORDS. disfluency, hireability, fillers, repetitions

BRIEF. Examining disfluencies in job interview settings.

ABSTRACT. Disfluencies have been studied in many contexts including laboratory settings where disfluencies are produced in noticeably scripted language, or conversational settings where disfluencies are produced in unscripted language. However, there are no studies on the specific impact of disfluencies on hiring decisions in an interview setting. Our overarching question is: can interview success be predicted by disfluency rates? Participants listened to one of three job interview audios with varying rates of disfluency (fluent, regular disfluent, high disfluent). Upon listening to the audio, participants were asked to rate (1) how likely they were to hire the candidate and (2) how effective of a communicator the candidate was. Data analysis from a one-way ANOVA revealed participants rated candidates in the fluent and regular disfluent groups equally in (1) hireability and (2) effective communication. For both hireability and effective communication, fluent was rated significantly higher than the high disfluent group ($p < .001$), and the regular disfluent group was rated significantly higher than the high disfluent group ($p < .01$). This suggests speakers who use above-average disfluency rates are less likely to be hired than speakers who use lower levels of disfluencies. This may be because listeners are not accustomed to high disfluency rates, so they may be more sensitive to speech including them.

INTRODUCTION.

Being an effective communicator is a vital skill in today's professional world. Communication is something that not only affects performance at a job but is taken into consideration at a job interview. Resources from the U.S. Department of Labor and the National Institute of Health provide tips on how to have successful job interviews, including ways to communicate during an interview, but there is an important aspect of communication that is left out which is known as fluency. There are two types of language regarding fluency: disfluent and fluent. Fluent language does not contain pauses, stutters, or interruptions. The opposite of this is disfluent language which has these disturbances.

Disfluencies are a natural part of language and are processed by our brains as such [1] This study aimed to discover what mechanisms listeners use to process disfluencies, but more specifically, it focused on fillers. The study found that listeners can decipher some disfluent sentences the same as they would a fluent sentence, depending on the location of a disfluency listeners may be more sensitive to a disfluent sentence than a fluent sentence. A different study investigated how repetitions are formulated by producers of disfluencies. This study found that repetitions appear when producers are having difficulty recalling language [2]. Fox Tree found that disfluencies can be as common as 6 per 100 words for the average producer overall [3].

Since disfluencies are a natural part of language it is important to understand how they affect language. Studies have shown disfluencies increase memory retention, particularly fillers such as “uh” and “um”. This contrast between the benefits and drawbacks of increased disfluency rates provides evidence that disfluencies may be an important factor in overall communication. Different disfluencies are indicators of different issues that may arise in language production [4]. For instance, fillers are the result of issues in message-level planning [5]. This is one reason why fillers do occur, but fillers might not be present when a speaker runs into a grammatical or phonological error in their speech production.

Job Interview settings are a nuance in disfluency research because it is scripted language catered to a particular setting, but it is still natural language. The fact that listeners know this may cause them to have higher expectations of speakers and be less lenient in the number of disfluencies they deem acceptable. We believe that the fluency used in language can affect the success of a candidate in a job interview. In this study, we examined how the rate of disfluencies impacts the likelihood of a speaker receiving a job as well as the effects it has on how well people perceive communication ability. We hypothesize that as disfluencies increase communication scores as well as hireabilty scores will decrease. We expect this since listeners are more sensitive to disfluent language than fluent language [1]. We think that listeners may be more sensitive to higher disfluency groups than they will be to the lower disfluency groups causing an inverse, not proportional, relationship between disfluency and interview success. We think that although the fluent group will receive higher scores than the other two experimental groups, there will be a smaller gap in scores between the fluent and regular disfluent group than the gap between the regular disfluent group and the high disfluent group. We believe because of the abnormal disfluent rate in the high disfluent group that the communication ratings and hireability scores will be significantly lower in the high disfluent than the other two groups. We want to further explore this topic in this study and others that follow it.

MATERIALS AND METHODS.

Participants.

30 Participants were recruited from Prolific, a crowdsourcing website. Inclusion criteria were that participants must be at least 18 years of age, a current U.S. resident, and a native speaker of English. Throughout the experiment 9 additional participants were excluded from data collection: leaving 30 participants for data collection with 10 in each experimental group. Reasons for exclusion included, spending too much time on a task, spending too little time on a task, using language processing software or robots to answer questions, failing 3 or more attention checks, awareness of the study goals, and not completing the survey. Upon exclusion of participants, new participants were added to the experiment until each group had 10 participants that met the inclusion criteria. All participants were given informed consent at the beginning of the survey and were able to leave at any time.

Script creation part 1.

The interview questions were created using questions from the Vanderbilt CALL interview process, and the Harvard Business Review article on the 10 most common interview questions. The questions were created so that they seemed familiar to participants, so participants were not confused by the content of the interview. We also wanted to make the questions general enough so that we could create long neutral responses so that the content of the response did not overshadow

the disfluency rates. The stimuli for the response were created by using common disfluent locations. Common disfluent locations were defined by findings from Alice in um-derland. They were paired with audio that was taken from mock interviews with research assistants. We coded all interview responses for repetitions, false starts the use of like, fillers, and silent pauses using a text-to-speech software called Notta that allowed us to transcribe WAV files to text, from which we could manually add the disfluencies.

Script Creation Part 2.

The other thing the research assistant's responses helped us with was creating the content of the interview. Some responses The CALL manager and we deemed neutral, so we took some of those responses and tried to replicate similar responses to questions. These two tools were combined in unison to create a framework for the Script of the project. The high disfluent script was made first for responses to all questions. Listeners may be more sensitive to disfluencies being placed in abnormal places, so keeping the disfluency locations consistent in the high disfluent group and the regular disfluent group was important. To keep these locations consistent half the disfluencies that were in the high disfluent group (12) were deleted in the regular disfluent group (6). The fluent group contained 0 disfluencies in the responses. Once the scripts were finalized the next task to complete was recording the audio for the scripts. The high disfluent group was the only group that had its script recorded. The other experimental groups were created by editing the audio from the high disfluent group to match the disfluency rates in the other two conditions. This was done to keep the prosodic features of all three experimental groups the same across all three conditions. Prosodic features entail how a speaker's voice sounds while articulating their speech. Parts of language like pitch or vocal intensity could be examples of prosodic features and some listeners may have different responses to differing prosodic features. By keeping one audio there is no difference in prosodic features from condition to condition. The deletion process was done through a software named PRAAT, which was cross-checked by two members of the team The platform used to display the Stimuli was REDCap. REDCap is an online survey tool that we used to create our 3 separate conditions. The survey included a consent form, a demographics portion, an interview portion, and post-survey questions.

Procedure.

To gather data on the three conditions being studied three separate surveys were sent out on Prolific; each survey contained one of three experimental groups which were fluent, regular disfluent, or high disfluent. A participant in any of the three surveys proceeded to take the survey on REDCap after signing up for the task. The survey was preceded with a consent form stating what the participant may have to do, and an option to leave the study at any time they wished. The next portion of the survey consisted of a demographic portion. All questions were optional, and no data was analyzed on the answers to these questions. In the next part of the Survey, participants were told to act as a hiring manager who needs to make judgment decisions for a company. The interview portion consisted of 5 different audio recordings that were responses to the 5 interview questions below. All recordings contained the questions below and the question was also displayed on the screen while the participants were listening to an audio. Each audio had its page and once the page was completed participants did not have an opportunity to return to a previous audio.

1. "Name two strengths and one weakness that you have."
2. "Do you prefer working independently or as a team why?"
3. "When you are balancing multiple projects how do you keep yourself organized?"
4. "Can you tell us about a time you received constructive criticism what was it? How did you respond to it? And what did you learn from the experience?"
5. "How do you deal with pressure or stressful situations?"

After hearing the audio, participants were asked to answer an attention-check question about the audio that they previously listened to. If a participant failed 3 or more attention checks, they were replaced with a different participant and their data was not included in the data analysis. The audios between the experimental groups were the same besides the disfluency rates that were present in the recordings of each group. The fluent group contained 0 disfluencies in any of the audios, the regular disfluent group contained 6 disfluencies per 100 words for each of the audios the high disfluencies group had 12 disfluencies per 100 words. After the interview, participants took the post-interview survey where they were asked to answer these 3 questions

1. How likely are you to hire this candidate?
2. How effective of a communicator was this candidate?
3. Please explain your hiring decision. Did anything stand out to you? If so, what? Why? Make sure to include advice for the applicant if you have any.

The first 2 questions were placed on a Likert scale from 1-7 with 1-2 being not likely or poor communicator 3-5 being somewhat likely or somewhat good of a communicator and 6-7 being very likely or a very good communicator. Upon completion of the third question, participants submitted the survey in its entirety.

Two one-way ANOVA tests were used to analyze the data from questions one (Name two strengths and one weakness that you have) and two (Do you prefer working independently or as a team why) from the post-interview to help answer the two questions. How does disfluency rate affect listeners' perceptions of communicative ability in disfluent speech and how do disfluency rates affect the likelihood of a candidate's being hired for a job

Hire Likelihood.

The one-way ANOVA and post-hoc analyses of Fig 1. indicate strong significant differences in hire likelihood between the fluent group and high disfluent group (p<.0018.56e-06; Figure 1). This means participants were significantly more likely to hire a candidate if they listened

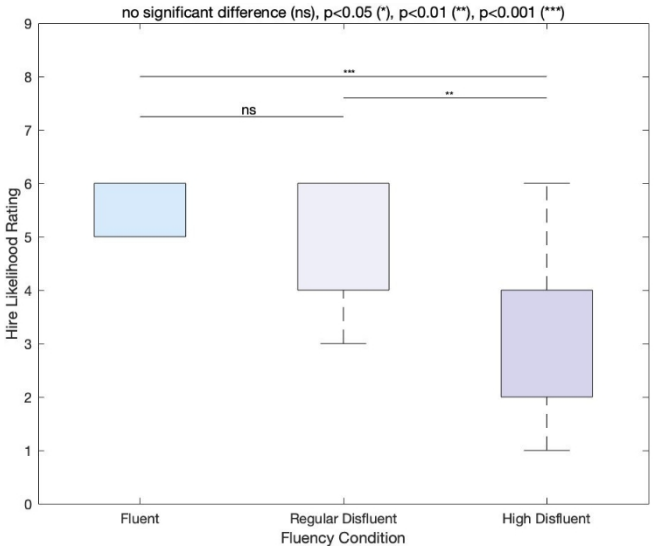

Figure 1. Hire Likelihood by condition. Hire likelihood is on y axis experimental groups on x axis scaled 1-7 High disfluent and regular disfluent scored significantly higher than fluent in communication rating.

to the fluent audio vs. the high disfluent audio. Subsequently, a significant difference was also noted between the regular disfluent Group and high disfluent Group ($p = 0.0015$). This shows the same trend seen in the Fluent group vs. high disfluent group except the margin isn't as wide. No significant differences were found between the disfluent group and regular disfluent group ($p = .1$).

Communication Rating.

The one-way ANOVA and posthoc analyses indicate strong significant difference in communication rating between the fluent group and high disfluent group ($p<.001$; Fig. 2). This means that participants in the fluent group rated the fluent interviewee as a significantly better communicator than the interviewee in the high disfluent group. We also observed a significant difference between the regular disfluent group and the high disfluent group ($p=.01$). No significant differences were found between the fluent group and the regular disfluent group ($p = .6$).

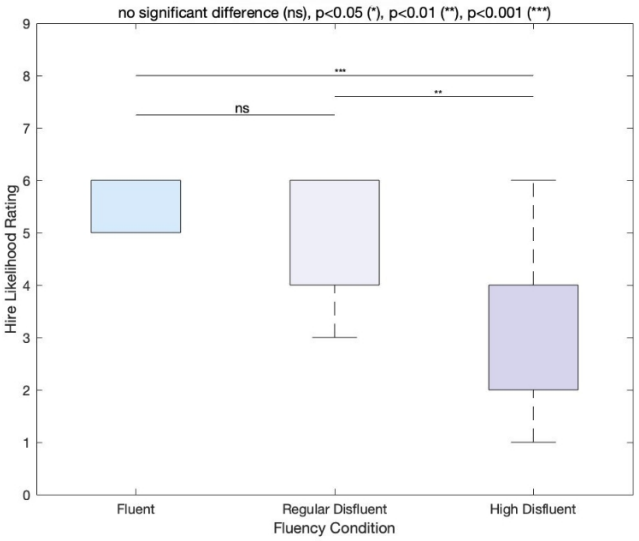

Figure 2. Communication rating by condition. Communication rating scores on y axis experimental groups on x axis scaled 1-7 High disfluent and regular disfluent scored significantly higher than fluent in communication rating.

DISCUSSION.

The results from question 1 do not support our first hypothesis because we found that the participants in the regular disfluent group were just as likely to hire a candidate as the participants in the fluent group. Similarly for question 2, how effective of a communicator is this applicant, participants in the regular disfluent group rated their candidate as highly as participants in the fluent group rated their candidate. In both conditions, participants in the high disfluent group had the lowest scores among all three groups and were significantly different than the other two conditions with a strong significant difference from the ratings of the fluent condition. There was no significant difference between the scores of the fluent and regular disfluent groups for either question. This might be because the regular disfluent group is what is naturally seen in language, so listeners are not as likely to have a negative response to a candidate that has a regular rate of disfluencies. Another reason this may have occurred is that listeners can be more sensitive to disfluent language than they are to fluent language. Meaning that disfluent language captures a listener's attention more so than fluent language because of the disfluencies that are present [1]. The rate of disfluencies used in the high disfluent group may have been exceedingly high and therefore could have made the listeners focus on the disfluencies more often than they would in the regular 'disfluent group with the average rate of disfluencies. This, along with negative traits that are associated with high disfluency rates, are a possible explanation for why participants in the high disfluent group gave the candidate a poor hireability score and rated them poor candidates.

The metric of perceived communication was used because of the varying factors that go into communication. Many prosodic features that impact communication were accounted for due to how the audios were recorded, but there are still components of being a good communicator that weren't accounted for. For instance, memory wasn't something that was considered in this experiment. A speaker's ability to help a listener remember what they said may be just as important as the listener's perception of how well they said it. On this premise, we believe that since disfluencies help improve memory retention [6] then having a natural rate of disfluencies may be the best way to become an effective communicator. This is because it helps improve memory retention and listeners did not rate them lower in hireability or rate them as a worse communicator than the participants did for the fluent candidate.

One limitation of this study was that there were only 10 participants in each experimental group. This was a pilot study that wasn't adequately powered ($N=30$) so there weren't enough participants to be certain about the results found from the experiment. A better-powered study with more participants in each experimental group will increase the accuracy of the results, the validity of the claims made, and the project's generalization.

A consideration for the future steps of this project is adding a two-way ANOVA or a mixed effects model to have a more accurate representation of the data. The two questions individually may not be the best way to display interview success in general, and even though there are hypotheses surrounding them the overarching question in this experiment is how interview success changes with increased rates of disfluency. So, mixing the two questions in a different statistical analysis may produce a more accurate representation of the data which in turn may answer the question more directly. One aspect of the post-interview that we should observe closely in the future is the third question in the post-interview. Some responses in the post-interview may help to alter the content of the response in the stimuli to become more neutral. Analyzing these responses may help to create new stimuli such as a job description or job resume to give participants more context before evaluating a candidate. This may also help tp change the responses so that factors other than disfluency have minimal impact on the responses of participants. One improvement that will be made to the third question in the next study is adding a word count minimum on the box where participants give their answers so we can receive insightful and helpful feedback rather than just adding a few words to complete the experiment. A final limitation of the experiment was the platform that we used. Online testing can result in the use of robots to generate responses. We made efforts to remove all bots from the data analysis but there isn't a guarantee that all were removed from the study. One way to circumvent this is by making participants take the test in person. This would ensure that all responses are genuine and stem from a human participant and not a robot.

In closing using disfluent language at a rate that is typical to a conversation may allow a speaker to be most successful in becoming an effective communicator and earning an um job.

ACKNOWLEDGMENTS.

I would like to thank Dr. Menton Deweese for supporting me through this project, Dr. Duane G. Watson for allowing me to do research in his lab, the CALL lab for being so welcoming of me, the SSMV for giving me this opportunity, and last but not least Doriana Lactignola for mentoring me throughout the entirety of this project, and pouring her time and effort into me and this project so that it could be it's absolute best.

REFERENCES

1. F. Ferreira, K. G. D. Bailey, Disfluencies and human language comprehension. *Trends in Cognitive Sciences* **8**, 231–237 (2004).
2. Z. Harmon, V. Kapatsinski, A theory of repetition and retrieval in language production. *Psychological Review* **128**, 1112–1144 (2021).
3. J. E. F. Tree, The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language* **34**, 709–738 (1995).
4. H. Clark, Using uh and um in spontaneous speaking. *Cognition* **84**, 73–111 (2002).
5. S. H. Fraundorf, D. G. Watson, Alice's adventures in *um* -derland: psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience* **29**, 1083–1096 (2014).
6. S. H. Fraundorf, D. G. Watson, The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language* **65**, 161–175 (2011).
7. J. E. Fox Tree, Interpreting Pauses and Ums at Turn Exchanges. *Discourse Processes* **34**, 37–55 (2002).
8. M. Corley, O. W. Stewart, Hesitation Disfluencies in Spontaneous Speech: The Meaning of *um*. *Language and Linguist. Compass* **2**, 589–602 (2008).
9. E. Diachek, S. Brown-Schmidt, The effect of disfluency on memory for what was said. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **49**, 1306–1324 (2023).
10. M. Susca, E. C. Healey, Listener perceptions along a fluency disfluency continuum: A phenomenological analysis. *Journal of Fluency Disorders* **27**, 135–161 (2002).
11. M. Corley, L. J. MacGregor, D. I. Donaldson, It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition* **105**, 658–668 (2007).

Amare Allen is a student at Martin Luther King Jr. Magnet High School in Nashville Tennessee he participated in a research internship through the School for Science and Math at Vanderbilt