Disfluencies in spontaneous speech in easy and adverse communicative situations: the effect of age

Linda Taschenberger, Outi Tuomainen, and Valerie Hazan
Department of Speech Hearing and Phonetic Sciences, UCL, UK

Abstract

Disfluencies are a pervasive feature of speech communication. Their function in communication is still widely discussed with some proposing that their usage might aid understanding. Accordingly, talkers may produce more disfluencies when conversing in adverse communicative situations, e.g. in background noise. Moreover, increasing age may have an effect on disfluency use as older adults report particular difficulties when communicating in adverse conditions. In this study, we elicited spontaneous speech via a problem-solving task from four different age groups (19-76 years old) to investigate the effect of energetic and informational maskers on the use of filled pauses (FPs), and its interaction with age. Measures of disfluency rates, effort ratings, and communication efficiency were obtained. Results show that, against our predictions, FP usage may decrease in adverse conditions. Moreover, age does not play a great role in adults with normal hearing. The results indicate that individuals differ greatly in their disfluency adaptations, utilising different strategies to overcome challenging communicative situations.

1. Introduction

Spontaneous communication is marked by its speech flow interruptions (Bortfeld et al., 2001). Two kinds of pauses are frequently examined in disfluency research: silent pauses, and filled pauses. Silent pauses are periods of non-articulation from a talker. Filled pauses, on the other hand, are periods of articulation of non-lexical content (Clark & Fox Tree, 2002), e.g. 'uh' and 'uhm' in English.

The effects of filled pauses (FPs) are widely discussed, with some proposing that they are used to buy the speaker time in order to plan their utterances (e.g. Jucker, 2015; Tottie, 2014) or hold the floor in conversation (Fox Tree & Clark, 1997; Shriberg, 1994). Others believe that FPs may prepare listeners for unexpected words and thereby actually aid understanding (Fox Tree, 1995; Arnold et al., 2003). Bortfeld et al. (2001) suggest that disfluencies may be affected by cognitive, social, and situational aspects. Shriberg (2001: 156) argues along similar lines stating that they "are related to the speaking environment in which they arise". Many have found an increase in disfluency (DF) use as task difficulty increases (e.g. Levin, Silverman & Ford, 1967). Some studies have also shown that talkers become more disfluent when they are speaking in background noise (Southwood & Dagenais, 2001). Furthermore, there are well-documented changes in both speech production and perception with increasing age. Some find that older talkers generate more disfluencies compared to younger adults (Bortfeld et al., 2001). Older adults (OAs) also often report having particular difficulty communicating in challenging situations (e.g. in noise). However, this interaction is not yet well explored; it is therefore not known whether OAs become more disfluent than younger adults (YAs) when communicating in adverse listening conditions.

The aim of the current study is therefore twofold: firstly, to explore the effect of filled pauses in the context of adverse listening conditions. Secondly, to investigate the role age plays in disfluency use. To investigate the impact of listening difficulty, we manipulated the type of interaction during completion of the same task: in the presence of 'energetic' (EM) & 'informational' masking (IM). It has been well established that the type of background noise can differentially affect listeners, with IM causing greater interference than EM (Rudner et al., 2012). We therefore recorded spoken interactions taking place (a) in quiet with no interference present, (b) in noise with no informational content (EM), and (c) in background speech (IM).

To elicit spontaneous interactive speech, we used a problem-solving 'spot-the difference' picture task (diapix, van Engen et al., 2010) under these three different listening conditions. We measured rate of disfluencies, communication efficiency (i.e., time it took to find differences), and self-rated listening effort and concentration. The research questions addressed are: 1) Does the rate of disfluencies differ across noise conditions? 2) Does the rate of disfluencies differ across age? 3) Do disfluencies have an effect on the communication efficiency and perceived effort & concentration of conversation?

We predict that the rate of FPs will be higher in the adverse conditions compared to the quiet condition. Furthermore, we predict that IM will result in higher use than EM as it is known to cause more interference. Regarding age, we predict that it will have an effect on frequency of FP use with older adults using more than younger adults in all conditions. Similarly to the age differences found in other speech strategies employed in noise, talkers

may adopt an increased use of FPs as a strategy to overcome the effect of background interference. Concerning the perceived effort of conversation, we do not have strong hypotheses: effort could either be rated lower and communication may be more efficient if an interlocutor uses more FPs as this may prepare listeners for unexpected words and thereby aid understanding. Alternatively, effort could be rated higher and communication could be less efficient if FPs are a marker of the talker struggling with the task at hand.

2. Method

2.1. Participants

54 monolingual native speakers of Standard Southern English participated in the study. They were aged between 19-26 years (Younger Adults, YA, N=20, 10 F, Mean age 21.8 years), 30-49 years (Middle Aged, MA, N=12, 8 F, Mean age 42.4), 50-64 (Older Middle Aged, OMA, N=10, 10 F, Mean age 61.8 years) and 65-76 years (Older Adults, OA, N=12, 10 F, Mean age 71.3 years). Participants were tested in pairs of the same sex within the same age band. All participants had normal hearing thresholds (<25 dB HL) across the 0.25-4 kHz range and reported no history of speech and language impairments or neurological trauma. All participants aged over 65 passed the Montreal Cognitive Assessment (MoCA) screening test.

2.2. Procedure

During the audio recordings, participants sat in separate acoustically-shielded rooms and communicated via headsets fitted with a cardioid microphone (Beyerdynamic DT297) whilst playing interactive Diapix games (Baker & Hazan, 2011) on a desktop PC. Participants were given different versions of the same picture scenes (e.g. Fig. 1) and told that they had 10 minutes to find the 12 differences between the pictures.

Each talker was recorded on a separate channel at a 44 100 Hz (16 bit) sampling rate using a Fireface audio interface and Audacity audio software. One of the talkers (designated 'Talker A') was told to lead the interactions. The other ('Talker B') was a more passive participant who mainly responded to queries by Talker A. All participants carried out both talker roles. After completion of every Diapix task, participants were asked to rate their effort and concentration (11-point Likert scale: "Did you have to put in a lot of effort to understand your partner?", 0=lots of effort, 10=no effort; "Did you have to concentrate very hard to understand your partner?", 0=concentrate hard, 10=not concentrate).

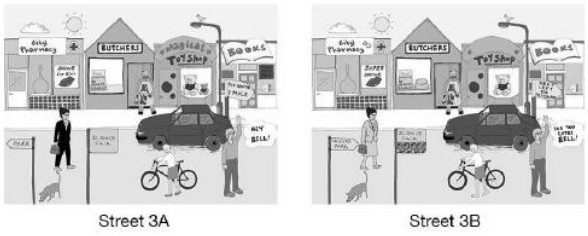

Figure 1: A DiapixUK picture pair

The picture task was carried out in four listening conditions affecting both participants: both speakers in i) quiet, ii) EM with no informational content ('speech-shaped' noise), iii) IM that is semantically related to the picture description task (i.e., talking about the same picture). The IM condition was a 3-talker masker consisting of a male, a female and a child speaker. The picture and noise condition orders were randomised. A fourth condition is not reported here. The apparatus and procedure were identical to Hazan et al. (2019). The data were collected as part of a wider protocol, but only relevant tasks are described here.

2.3. Data processing

All recordings were automatically transcribed using a speech recognition system by Speechmatics and then manually corrected in Praat (Boersma, 2011). Annotation of filled pauses was performed against the word level transcriptions by the first and second author using the spellings UH, UHM, UM, ER, ERM, UH. From the recordings we calculated measures that reflect i) Talker A's amount of FPs as a percentage of total utterances spoken (as recording durations were not consistent), ii) communication efficiency (i.e., time in seconds from start to finding 8th difference), iii) listening effort and concentration ratings for Talker B.

3. Results

3.1. Filled pauses

Overall, filled pauses usage reflected other reported findings with FPs taking up an average of 3.86% of all utterances (Clark, 1994). To establish whether FP use differed across noise conditions, a repeated measures ANOVA was carried out for the within-subject factor listening Condition (3: QUIET, EM, IM). In general, use of FPs was higher in the quiet condition (M=4.27%, SD=2.20) than in both of the adverse conditions (EM: M=3.67%, SD=1.90; IM: M=3.66%, SD=1.93). This main effect of Condition was significant ($F(3,159)=3.929$, $p<0.01$) against our hypothesis that disfluency use would be greater in adverse conditions. Post hoc analyses showed no difference across adverse conditions, but that this significance was driven by the differences to QUIET.

Age analyses were based on linear mixed-effects modelling using the lme function in the nlme package for R (Version 1.1.463). The best-fitting model for each individual analysis was chosen with hierarchical approaches, that is, adding one predictor at a time to a baseline model that includes no predictors other than the intercept. Condition (3) and Age (continuous; centred at the mean across age bands) were entered one by one as fixed effects and Participant as random effect. Likelihood ratio tests were used to determine which effects were needed in the model. Age showed a statistically significant interaction effect (p<0.01) in only the IM condition (see Fig. 2). Here, FP usage declined with increasing age.

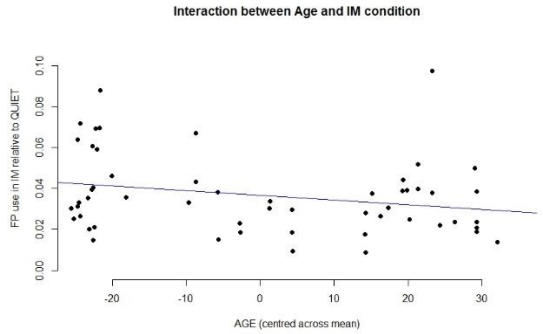

Figure 2: Interaction effect between Age and IM condition for FP use (as proportion of total utterances)

As these findings go against our predictions and all show high variance, we decided to investigate the individual differences in the data. We firstly calculated the percentage change relative to FP use in QUIET for all participants. This showed that participants greatly differed in their use of FPs overall and across conditions (see Fig. 3). Some increased their usage in adverse conditions, while others decreased it. We then calculated individual z-scores for each participant for each condition, with a z-score at ±1 SD taken as a meaningful difference, denoting that the individual was making marked changes in their FP production across conditions (i.e. outside the 15th-85th percentile range). This showed that in EM, 12 individuals differed from the population mean, five were below the 15th percentile and seven above the 85th percentile. This indicated a change of -50% to -100% and 39% to 113% change relative to QUIET, respectively. In IM, 15 individuals differed from the population mean, with seven found under the 15th percentile and eight over the 85th percentile. This indicated a change of -50% to -70% and 42% to 100%, respectively. It is also of interest to see if speakers were consistent in their FP strategy across adverse conditions. Five participants meaningfully increased their FP use in both adverse conditions. One individual meaningfully decreased FP rate in both adverse conditions. One individual decreased in EM and increased in IM. The rest of the participants only showed a significant change in FP frequency in one adverse condition.

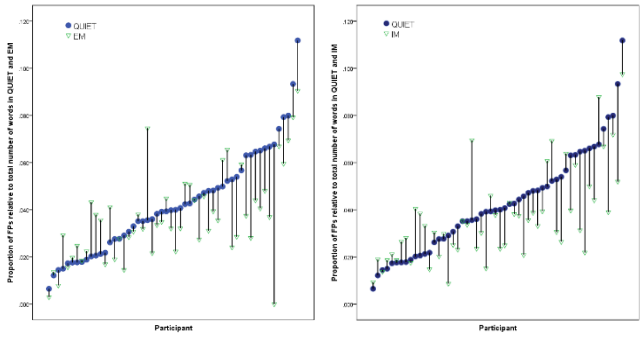

Figure 3: FP rate for each participant between QUIET and EM (left) and QUIET and IM (right)

3.2. Perceived effort & communication efficiency

In order to investigate whether ratings given immediately after completing Diapix were associated with FP use, we carried out bivariate Pearson's correlation calculations for mean ratings as a function of partner's FP use. These analyses were done for two ratings (described in 2.3) evaluating both effort and concentration. For Effort ratings, the only significant correlation at the p<0.01 level was for the IM condition, with higher effort ratings in higher FP use instances (r=-0.34). With regards to communicative efficiency, we correlated whether Talker A's FP use had an effect on the time it took to find eight differences in the pictures. This factor only played a minor role in QUIET (p=0.036, r^2=0.08). None of the adverse conditions were influenced by FP use here. Overall, these correlation analyses showed only weak relationships with most going in the opposite direction of our predictions.

4. Discussion

It has often been found that talkers increase their disfluency use in conversational speech as task difficulty (Levin, Silverman & Ford, 1967; Taylor, 1969) or background noise increases (Jou & Harris, 1992; Southwood & Dagenais, 2001) or with increasing age (Bortfeld et al., 2001). In the current study which recorded conversational speech with communicative intent in easy and adverse speaking conditions in speakers spanning the adult age range, all with normal hearing thresholds up to 4 kHz, we could not replicate these findings. Overall, use of filled pauses decreased in the more challenging situations. Additionally, age played only a minor role. However, our data set showed large variance: analyses of individual differences suggest that speakers adopt differing strategies in their speech adaptations. When communication becomes

effortful, talkers need to make various adjustments to their speech production to aid listeners' understanding, for example, by speaking more slowly and reducing complexity of their utterances (Gagné, 1994). There are large individual differences in how effectively clear speech is produced (e.g. Hazan et al. 2018). This may extend to disfluency use, too. Background interference may not affect everyone to the same degree and result in the same production strategies. It has indeed been widely noted that disfluency rates vary between corpora (e.g. Branigan et al., 1999; Lickley, 2001) with different overall dialogue tasks and speaker roles greatly affecting the way disfluencies are produced.

Importantly, we solely investigated the use of filled pauses. It may be other types of disfluencies that increase in more challenging situations (e.g. hesitations or repetitions). Complementary analyses on silent pauses will therefore be carried out and be presented at the conference. It may be the case that the less filled pauses are used in adverse conditions, the more silent pauses are utilised. This additional information will tell us more about the strategies adopted in adverse speaking conditions.

However, our lack of a significant age effect might also be due to the fact that a decline in sensory acuity (e.g. hearing ability) may be a contributing factor. Tuomainen & Hazan (2017) found a significant relationship between hearing thresholds and disfluency usage with poorer hearing resulting in more disfluent conversations. As our participant group all had hearing thresholds under 25dB, this might have eliminated the conflating age effect found in some other studies.

Acknowledgements

This work was supported by the Economic and Social Research Council [grant number ES/P002803/1].

References

Arnold, J. E., M. Fagnano, M.K. Tanenhaus, 2003. Disfluencies signal theee, um, new information. *J. Psycholinguist. Res*. 32 (1), 25–36.

Baker, R., & V. Hazan. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43, 761–770.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. Glot International 5(9). 341–345.

Bortfeld, H., S.D. Leon, J. Bloom, M.F. Schober, & S. Brennan. 2001. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. Language and Speech, 44, 123–147.

Branigan, H., R. Lickley, & D. McKelvie. 1999. Non-linguistic influences on rates of disfluency in spontaneous speech. *Proceedings of the 14th International Conference of Phonetic Sciences.*

Clark, H. 1994. Managing problems in speaking. *Speech Communication*, 15, 243–250.

Clark, H. H. & J. E. Fox Tree. 2002. Using uh and um in spontaneous speaking. *Cognition* 84(1):73–111.

Cloud transcription service. https://www.speechmatics.com

Fox Tree, J. E. 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34(6), 709–738.

Fox Tree, J.E. & H. H. Clark. 1997. Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition* 62 (2). 151–167.

Gagné, J.P., V.M. Masterson, K. G. Munhall, N. Bilida, & C. Querengesser (1994). Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *J. Acad. Rehab. Audiol.* 27, 135–158.

Hazan, V., O. Tuomainen, J. Kim, C. Davis, B. Sheffield, & D. Brungart. (2018). Clear speech adaptations in spontaneous speech produced by young and older adults. *The Journal of the Acoustical Society of America*, *144*(3), 1331-1346.

Hazan, V., O. Tuomainen & L. Taschenberger. 2019. Subjective evaluation of communicative effort for younger and older adults in interactive tasks with energetic and informational masking. *Proceedings of the 19th International Congress of Phonetic Sciences*

Jucker, A. H. 2015. Uh and Um as Planners in the Corpus of Historical American English. Developments in English: Expanding Electronic Evidence, 162–7

Levin, H., I. Silverman & B. L. Ford. 1967. Hesitations in children's speech during explanation and description. *Journal of Verbal Learning & Verbal Behavior*, 6(4), 560–564.

Lickley, R. J. 2001. Dialogue moves and disfluency rates. *In DISS'01*, 93-96

Rudner, M., T. Lunner, T. Behrens, E.S. Thorén, J. Rönnberg. 2012. Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology* 23(8):577–589.

Southwood, M. H. & P. Dagenais. 2001. The role of attention in apraxic errors. *Clinical Linguistics and Phonetics,* 15, 113–116.

Shriberg, E. 1994. *Preliminaries to a Theory of Speech Disfluencies*. PhD thesis, University of California at Berkeley

Shriberg, E. 2001. To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association,* 31, 153–169.

Tottie, G. 2014. On the Use of Uh and Um in American English. *Functions of Language* 21(1):6–29.

Tuomainen, O. & V. Hazan. 2017. Disfluencies in spontaneous speech in younger and older adults in easy and difficult communicative situations. *Workshop on Challenges in Analysis and Processing of Spontaneous Speech.*

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., Bradlow, A. R. 2010. The Wildcat Corpus of Native- and Foreign-Accented English: communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech,* 53, 510-540.