Speech Communication 95 (2017) 16–27

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Profiling fluency: An analysis of individual variation in disfluencies in adult males

Kirsty McDougall[a,b,*], Martin Duckworth[1]

[a] Clare College, Cambridge CB2 1TL, United Kingdom
[b] School of Humanities, University of Hertfordshire, College Lane, Hatfield AL10 9AB, United Kingdom

ARTICLE INFO

Keywords:
Fluency behaviour
Disfluency features
TOFFA
Individual differences
Speaker-specificity

ABSTRACT

Individual variation in non-fluency behaviour in normally fluent (NF) adults, is investigated. Differences among speakers in the usage of a range of features such as filled and silent pauses, sound prolongations, repetition of phrases, words or part-words, and self-interruptions is explored in the spontaneous speech of 20 male speakers of Standard Southern British English from the *DyViS* database. The speech analysed is semi-spontaneous, and taken from a simulated police interview task. A taxonomy of fluency features for forensic analysis (TOFFA) was applied to this speech data. The rate of occurrence of each feature per 100 syllables is calculated for each speaker. Results show that individuals vary considerably in the rates of these fluency features occurring in their speech and that between-speaker differences are present in the types of features speakers produce. Implications of the significance of these findings for forensic phonetics are discussed.

1. Introduction

Speakers interrupt the flow of their speech in different ways. Speech features relating to fluency such as filled and silent pauses, sound prolongations, repetitions and self-interruptions have been examined in a range of studies of fluency, with particular interest in the role of such phenomena in speech planning and productions and their relevance to the comprehension of speech (e.g. Blankenship and Kay, 1964; Brennan and Schober, 2001; Corley et al., 2007; Fraundorf and Watson, 2011; Goldman Eisler, 1961; MacGregor et al., 2010; Shriberg, 2001). Fluency phenomena exhibit variation between speakers, yet the speaker-specificity of such features has received little attention in phonetic research. The present study investigates the extent of individual variation exhibited by such fluency disruptions to a speaker's flow of speech and considers whether such variation might contribute to a profile of speech phenomena which could assist in the forensic analysis of speech recordings.

Disfluencies have been extensively investigated in the speech of people who stutter (PWS) and some of the earliest taxonomies of disfluencies were produced by speech pathologists (e.g. Johnson, 1961; Johnson et al., 1963; Van Riper, 1973; Wingate, 1964). Van Riper (1973) comments that the features of a PWS's speech "is as unique as their fingerprints" (1973: 128). The possibility that fluency disruptions in PWS may manifest speaker-specific characteristics prompts the question of whether disfluency in NF speakers is also speaker-specific.

An examination of the literature yields relatively few studies in which the patterns of fluency phenomena of NF speakers are applied to the speech of PWS. There are however studies where definitions of phenomena found in the literature on stuttering are applied to NF speakers. Johnson et al. (1963) using data first presented in Johnson (1961) provides results of an analysis of the fluency features of 50 male and 50 female PWS and 50 male and 50 female NF speakers. The participants were asked to produce three monologues. Roberts et al. (2009) replicated the broad outline of this study using 25 NF adult male speakers. In their summary of the findings of Johnson (1961) and a number of other studies of fluency in NF speakers, Roberts et al. (2009) commented upon the diversity of methods of counting disfluencies in earlier studies. For example, the phenomena may be related to the frequency of occurrence per 100 words or 100 syllables but may not define what is counted as a word or a syllable. Roberts et al. counted interjections including filled pauses and utterances like 'well', 'like', 'you know' into this category. They also counted revisions, repetitions, prolongations and the use of 'excessive force in producing a sound' (2009: 425) which they termed a *block*. They did not count silent pauses. Roberts et al. relate the fluency phenomena to occurrence per 100 syllables which were defined as target (i.e. idealised phonological) syllables only. They report that individuals produce a range of fluency phenomena per 100 syllables yet even speakers with double the rate of other speakers 'still appear to be

* Corresponding author at: Clare College, University of Cambridge, Cambridge, CB2 1TL, United Kingdom.
E-mail address: kem37@cam.ac.uk (K. McDougall).
[1] Independent Researcher at: c/o Dr Kirsty McDougall, Clare College, Cambridge CB2 1TL, United Kingdom.

http://dx.doi.org/10.1016/j.specom.2017.10.001
Received 28 May 2017; Received in revised form 8 October 2017; Accepted 9 October 2017
Available online 10 October 2017
0167-6393/

speaking well' (2009: 424). Roberts et al. do not comment on the speech task they employed, possibly because they were guided by other studies of speech on PWS in which monologues had been used. Monologues bear limited relationship to speech produced in an interaction. Those studying the potential functionality of fluency phenomena in NF speech have used conversational speech. For example, both Eklund (2004) and Shriberg (2001) use speech derived from a corpus of telephone calls made to organise travel arrangements. Although Roberts et al. (2009) examined the speech of NF speakers they defined some of the phenomena they examined in terms of phenomena which might be found in PWS. Allwood et al. (1992) and Gilquin and De Cock (2011) suggest that rather than assuming that all non-fluencies demonstrate a lack of competence they should be considered a natural phenomenon of spontaneous human speech contributing both to the perception of fluency and effective interaction between speakers. Fox Tree (2001) for example, has demonstrated that the use of filled pauses helps listeners to understand a word following a filled pause.

1.1. What is fluent speech?

Listeners do not normally notice when NF speakers are not perfectly fluent, therefore what a listener imagines to be perfectly fluent is likely to contain several disfluencies. Perfect fluency does not occur in NF speakers because disfluencies are necessary. Filled and silent pauses, repetitions, prolongations and changes to the topic of speech all occur in NF speech, but how often do these phenomena occur? In some ways, it may seem counter-intuitive to regard phonological lengthening as a disfluency because it plays a role in speech prosody. However, this study seeks to determine how much segment-lengthening NF speakers use rather than assuming what is and is not a correct amount of prolongation. Determining the range and variability of disfluencies in NF speakers will enable the speaker-specificity of such behaviours to be explored and provide data against which the speech of other speakers may be compared.

1.2. Why might fluency disruptions be speaker-specific?

The topic, the speed at which a speaker can process information, and the effect of the topic and speech context upon a speaker's ability to formulate and execute a response may all affect the fluency of speech. Most speech is unlikely to be wholly pre-planned and there is no reason to assume that all speakers will plan or produce speech in the same way. Speakers can employ different strategies and a given speaker might tend to use some strategies rather than others. For example, there are both psychological and prosodic explanations for the occurrence of filled and silent pauses (e.g. Brennan and Schober, 2001; Corley et al., 2007; Fraundorf and Watson, 2011; Goldman-Eisler, 1968). The frequency and location of the pauses may therefore be influenced by the individual speaker's response to psycho- or socio-linguistic demands. Some authors (e.g. Kjellmer, 2003) consider that filled pauses are equivalent to words and unfilled pauses may serve a function for both the speaker and the listener. Other breaks in fluency such as repetition, prolongation and self-generated interruption might also reflect the speech planning and execution process and therefore can similarly be expected to exhibit individual variation.

Individual variation in fluency disruptions has received little attention in the literature. Künzel (1997: 51) observed in relation to filled pauses that '[i]ndividuals tend to be quite consistent in using "their" respective personal variant of the hesitation'. Hughes et al. (2016) cite this quotation and predict that '[filled pauses will] show relatively little within-speaker variability' (2016: 101). Hughes et al. undertake an acoustic analysis of filled pauses in 86 English-speaking males using a likelihood ratio framework. Hughes et al. suggest that focussing on the acoustic analysis of a single type of disfluency may be relatively less difficult than the broader analysis of disfluencies explored in the present study. Their results show that the acoustic analysis of filled pauses does indeed demonstrate intra-speaker consistency and has the potential to contribute to discriminating between speakers. Also with forensic motivations, Schiel and Heinrich (2015) and Braun and Rosin (2015) both examine a range of disfluency phenomena. Schiel and Heinrich report on changes in the occurrence of six disfluency phenomena in 150 German male and female speakers following the consumption of alcohol. These authors concentrate on the group changes in the occurrence of these disfluencies rather than on speaker-specific patterns. They allude to the 'idiosyncratic behaviour' (2015: 30) of some speakers though this relates to idiosyncratic effects of alcohol on different speakers. With forensic discrimination between speakers in mind, Braun and Rosin (2015) study the occurrence and distribution of seven types of disfluency phenomena in three monologue tasks undertaken by ten female German speakers. They examine three different types of filled pause and four different types of segmental prolongation. Unfilled pauses, repetitions and false starts were excluded from their taxonomy. Unlike most other studies Braun and Rosin extract the frequency of occurrence of per minute rather than per number of syllable or words. They report variation both in the overall rates of disfluency in their ten subjects across the three tasks, and in the frequency of use of the different types of disfluency examined in their study. They comment that speakers tend to use four or five different types of disfluencies and that individuals may focus on the use of one or two of these.

With the exception of Braun and Rosin (2015), studies of fluency behaviour in NF speakers have not included the degree of speaker-specificity of different types of disfluency features as a research focus. However, various researchers have commented on individual differences in the range, pattern and frequency of occurrence of these features. A brief review of some of these findings is outlined below, bearing in mind that numerical comparisons across studies are difficult because different methods of eliciting speech are used. Further, there are differences both in the disfluency taxonomies used and in the ways in which each study compares the occurrence of disfluencies against the whole speech sample (e.g. per 100 syllables, per 100 words, or per minute of speech). Shriberg (2001) working within a conversational interaction framework excluded unfilled pauses and prolongations in her examination of corpora of different types of English telephone interactions. Eklund (2004), working within a similar framework, analysed several corpora of Swedish telephone conversations. He included both unfilled pauses and prolongations in his taxonomy. Both Shriberg and Eklund relate occurrence of disfluencies to the total number of words rather than the number of syllables. Roberts et al. (2009) examined the occurrences of disfluencies in three different monologues from 25 English speakers. The authors' aim was to compare the results with data from people who stutter and the taxonomy of disfluencies, which excluded unfilled pauses, was influenced by this perspective.

Table 1 below permits comparison of the mean, lowest and highest disfluency rate results. All unfilled pause data, where collected, has been excluded. To simplify the comparisons, the rates are given as disfluencies per 100 words (or per minute in the case of Braun and Rosin, 2015). Where the original data were expressed in number of disfluencies per 100 syllables the rate has been converted to per 100 words using a conversion formula in which the number of syllables is multiplied by 0.7143 (Andrews and Ingham, 1971).

Although the results shown in Table 1 are not directly comparable due to the different means used for data elicitation, the different taxonomies applied, and the different means of expressing the amount of disfluency, the studies all demonstrate that different speakers produce different amounts of disfluency. The present study aims to drill deeper into these individual differences, by examining speaker-specific variation in the profile of disfluency types produced by individual speakers. Eklund (2004) and Roberts et al. (2009) both comment on the wide variety of amount and types of disfluency used by different speakers. Shriberg (2001) observes that speaker-specific strategies are present in the types of fluency phenomena NF speakers use, depending on the cognitive demands of the task. For example, she found that