

**Citation for published version:**

Kirsty McDougall, and Martin Duckworth, 'Profiling fluency: An analysis of individual variation in disfluencies in adult males', *Speech Communication*, Vol. 95:16-27, December 2017.

**DOI:**

<https://doi.org/10.1016/j.specom.2017.10.001>

**Document Version:**

This is the Accepted Manuscript version.

The version in the University of Hertfordshire Research Archive may differ from the final published version.

**Copyright and Reuse:**

© 2017 Elsevier B. V.

This manuscript version is made available under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License CC BY NC-ND 4.0,

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

**Enquiries**

If you believe this document infringes copyright, please contact the Research & Scholarly Communications Team at [rsc@herts.ac.uk](mailto:rsc@herts.ac.uk)

# **Profiling Fluency: An Analysis of Individual Variation in Disfluencies in Adult Males**

Kirsty McDougall<sup>a,b</sup> and Martin Duckworth<sup>c</sup>

<sup>a</sup>University of Cambridge <sup>b</sup>University of Hertfordshire <sup>c</sup>Independent Researcher

Dr Kirsty McDougall (corresponding author)

Postal address 1:  
Clare College  
Cambridge CB2 1TL  
United Kingdom

Postal address 2:  
School of Humanities  
University of Hertfordshire  
College Lane  
Hatfield AL10 9AB  
United Kingdom

email: kem37@cam.ac.uk

Martin Duckworth

Postal address:  
c/o Dr K. McDougall  
Clare College  
Cambridge CB2 1TL  
United Kingdom

email: martinsduckworth@gmail.com

# Profiling Fluency: An Analysis of Individual Variation in Disfluencies in Adult Males

## 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.

**KEYWORDS:** fluency behaviour, disfluency features, TOFFA, individual differences, speaker-specificity

## HIGHLIGHTS

- A detailed taxonomy of disfluency types (TOFFA) is described.
- Individual variation in a range of fluency features is observed.
- A consistency study demonstrates the challenges of identifying disfluencies.
- The significance of disfluencies for forensic speaker comparison is considered.

## 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 & Kay, 1964; Brennan & Schober, 2001; Corley, MacGregor, & Donaldson, 2007; Fraundorf & Watson, 2011; Goldman Eisler, 1961; MacGregor, Corley, & Donaldson, 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, Darley, & Spriestersbach, 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, Darley and Spriestersbach (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, Meltzer and Wilding (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 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, Nivre and Ahlsén (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 & Schober, 2001; Corley *et al.*, 2007; Fraundorf & 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, Wood and Foulkes (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 & Ingham, 1971).

Table 1. Overall disfluency rates produced by speakers (mean across the speakers, lowest rate for an individual, highest rate for an individual) in previous studies. The rates are per 100 words except for Braun and Rosin (2015) which are per minute

Study	Language and speech task	Mean rate	Lowest rate	Highest rate
Shriberg (2001)	English telephone conversations	6	1	18
Eklund (2004)	Swedish telephone conversations	6.4	1.6	16.1
Roberts <i>et al.</i> (2009)	English monologues	5.1	1	10
Braun and Rosin (2015)	German monologues	5.2 per min	4.5 per min	12.3 per min

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 certain speakers tended to repeat elements of speech, e.g. *all the-the tools*, while others were characterised by deleting elements of speech, e.g. *it's- I could get it where I work*. The present study will quantify the extent of individual variation in speakers' disfluency feature profiles, and explore the speaker-characterising potential of the profiles.

### 1.3 Why might a speaker-specific analysis of fluency features be useful in forensic speaker comparison?

The analysis of individuals' fluency has potential for application in forensic speaker comparison where the speech of an unknown speaker committing a crime and a suspect are compared to determine whether the same speaker is present on both recordings. A good deal of the literature on speaker-distinguishing properties of speech for forensic applications has focused on phonetic features which bear a direct relationship with a speaker's anatomy, for example mean fundamental frequency (pitch) which reflects the length and mass of a speaker's vocal folds (e.g. Hudson, de Jong, McDougall, Harrison, & Nolan, 2007; Künzel, 1989), or formant frequencies which reflect the dimensions and configuration of the vocal tract (see e.g. Foulkes & French, 2012 and references therein; Jessen, 2008). Investigating the speaker-distinguishing potential of fluency features focuses on a behavioural rather than an anatomical aspect of a speakers' performance. In addition, fluency phenomena are realized in the temporal domain and therefore relatively well-preserved in a poor recording where background noise and telephone transmission (with its reduced bandwidth) are typical.

Another reason for the value of this aspect of speech in forensic case work stems from the observation that disfluency is a normal, natural aspect of speech. As has been suggested above fluency features in NF speakers is not usually noticed by either speaker or listener. If this is the case, then speakers are likely to find it very difficult to modify deliberately the frequency and type of disfluencies they use.

The present study quantifies the individual variation in disfluencies occurring in a group of NF adult males. It aims to determine the range of usage of fluency phenomena by NF speakers and the extent to which the profiles of disfluencies used by speakers are speaker-specific.

## **2 TAXONOMY OF FLUENCY PHENOMENA AND TRANSCRIPTION CONVENTIONS USED**

For the present study, the following general definition of a 'fluency disruption' was adopted:

*any phenomenon originated by the speaker which changes the flow of the speaker's utterance.*

This definition is deliberately broad to reflect the issues about the nature of fluency raised in the introduction. The categorisation of the different types of fluency phenomena identified in this study were informed both by the work of others and by observations made in relation to the present dataset. A system for classifying the speech disfluencies in each recording was devised combining features used by those studying NF speech and by those examining stuttered speech. We propose that this taxonomy is called the Taxonomy of Fluency for Forensic Analysis, or TOFFA. The structure of TOFFA is outlined below.

### **2.1 General types**

The fluency phenomena in TOFFA break down into five groups:

1. **Unfilled pauses**
2. **Filled pauses**
3. **Repetitions**
4. **Prolongations**
5. **Interruptions**

Definitions and examples of each of these general types and the subcategories within are provided in Sections 2.2 – 2.6 following. Orthographic and occasionally phonetic transcriptions of examples of each disfluency type are included to assist in interpreting the coding. To ensure the disfluency codes are not confused with phonetic transcriptions the codes are formatted in bold and within square brackets. The text of the transcriptions in the examples below is given in italics.

## 2.2 Unfilled Pauses

**2.2.1 Definition:** a silence  $\geq 200$  msec occurring within a single speaker's turn. These are referred to as silent pauses by some authors.

**2.2.2 Types:** There are two types of unfilled pause, which are based on the definitions in Goldman-Eisler (1968: 13):

1. **[pg]** pauses at grammatical boundaries, e.g. *he came [pg] and I left shortly after*
2. **[po]** other pauses, i.e. those located elsewhere, e.g. *he came and I [po] left shortly after*

The above follow the examples provided by Goldman Eisler where the [pg] occurs before a conjunction and the [po] occurs 'in the middle ... of a phrase' (1968: 13). However, there may be other prosodic features such as a rising pitch on *I*. This might suggest that the following pause before *left* is a deliberate pause and is therefore grammatical pause. This is explored further in 4.2.

**2.2.3 Duration:** Goldman-Eisler (1968) proposed that an unfilled pause should last at least 250 msec to avoid purely mechanical artefacts. However other writers, e.g. Butterworth (1980), use 200 msec and, more radically, Kirsner, Dunn and Hird (2003) observe that different speakers (and even the same speakers on different occasions) may use different minimum pause lengths. They further noted that there may need to be a long and a short pause category for each speaker. The present study used a minimum of 200 msec. Using categories of long and short silent pauses was not possible because of the difficulty in determining what the speaker-specific boundaries for these terms should be.

### 2.2.4 Unfilled pauses at the start of a speaker's turn:

The time between the end of one speaker's utterance and the first vocalisation of the next speaker is not counted as an example of an unfilled pause. However, a pause following an initial filled pause is counted, e.g.:

Question: *what time did you arrive?*

Answer: [unfilled pause – not counted] **[er]** **[pg]** *I arrived soon after that.*

Here one filled pause **[er]** and one unfilled pause **[pg]** are counted.

## 2.3 Filled Pauses

**2.3.1 Definition:** a vowel sound, which may or may not be followed by a nasal.

**2.3.2 Types:** Two major types and one less frequent type. The three types and their codes are:

1. **[er]** vowel alone
2. **[erm]** vowel plus nasal (invariably /m/ in the present study)
3. **[fpo]** filled pause 'other' with a different or no vowel

A speaker might use a number of variants to the vowel and the nasal, but as long as there was a - roughly - mid-central vowel **[er]** and **[erm]** were used. **[fpo]**: Speakers occasionally use a



different vowel, e.g. [ɑ:] [u:] or omitted the vowel e.g. [m:]. These types of filled pauses are classified as filled pause ‘other’ and **[fpo]** is used.

**Note:** it is also common to refer to type 1 as *uh* and type 2 as *uhm* to avoid the possibility that readers might imagine that the *r* had to be pronounced. In this study *er* and *erm* are used.

**2.3.3 Durations:** There are arguments that short and long forms of filled pause have different functions, e.g. Kjellmer (2003) but the present taxonomy does not distinguish between long and short filled pauses.

**2.3.4 Sequences of filled and unfilled pauses during a speaker’s turn:** In sequences such as **[pg]** + **[er]** + **[pg]** between two utterances by the same speaker each of the pauses is counted separately, e.g. *I arrived soon after that [pg] [er] [pg] but he wasn’t there.* Here two unfilled pauses, both **[pg]**, and one filled pause **[er]** are counted.

## 2.4 Repetitions

**2.4.1 Definition:** a part of a word, whole word or phrase is repeated; there may be more than one iteration.

**Note:** Eklund (2004) includes repetitions in what he defines as repairs which also includes substitutions, insertions and deletions. This interpretation subsumes repetition into a different category. The present study, consistent with a number of other analyses, e.g. Shriberg (2001) Roberts *et al.* (2009), Schiel and Heinrich (2015), regard repetitions as a distinct type of disfluency.

**2.4.2 Types:** There are three types of repetition and one variant on these:

1. **[wrep]** whole word repetitions, e.g. *I-I [wrep] arrived later.*

There may be an immediate repetition as in the above example or a pause (filled or unfilled) or a sequence of pauses may intervene between iterations, e.g. *yes I-[po] [er] [po] I [wrep] walk there sometimes.* All such pauses are counted.

Elided phrases such as *it’s* for *it is* are counted as a single word hence *it’s-it’s [wrep] time to go* is coded as a word repetition rather than a phrase repetition.

Repeated words which might be regarded as rhetorical such as *yes-yes* are counted as whole word repetitions.

2. **[pwr]** part word repetitions, e.g. [tə.ten] for *t-ten***[pwr]**, or [s.sɪks] for *s-six***[pwr]**.

3. **[prep]** phrase repetitions, e.g. *I want to- want to***[prep]** *stay.*

4. **[mrep]** variant: multiple repetitions, e.g. [tə.tə.ten] *t-t-ten***[pwr]****[mrep]**; *I-I-I***[wrep]****[mrep]** *was there.*

**2.4.3 Note:** **[mrep]** is used as an *additional* code whenever more than one iteration of a repetition of any type occurs. Both the type of repetition and the multiple repetition type are coded. No more than one **[mrep]** code is used in a single series of repetitions regardless of the number of iterations. As with the other types of repetitions, filled or unfilled pauses or a combination of these might occur between repetitions. All of these pauses are counted.

## 2.5 Prolongations

**2.5.1 Definition:** one or more speech segments in a word are prolonged. Prolonged filled pauses are, however, not counted as a prolongation.

**2.5.2 Duration:** In all cases a prolongation means that a segment lasts for  $\geq 200$  msec regardless of whether the prolongation might have been motivated by the prosody of an utterance (see also 4.4). In the orthographic transcription the code is placed as close to the event as possible to indicate which segment has been prolonged.

**Note:** Eklund's definition of a prolongation is somewhat subjective in that he describes it as 'phones that are longer than should be expected in normal-paced, fluent, speech' (2004: 163). Roberts *et al.* (2009: 425) employ a similar subjective definition 'Any sound ...judged longer than normal' with the added criterion: 'if a stuttering client said it that way, I would count it as a prolongation'. Schiel and Heinrich (2015: 25) also use a subjective judgement of prolongation or 'unusual phone lengthening' which is identified as 'judged by the listener'. Both the present study and Braun and Rosin (2015) use a specific duration to identify a prolongation. Braun and Rosin use 300 msec (given by Braun in answer to a question at the ICPHS conference); the present study uses 200 msec.

**2.5.3. Types:** There are three types of prolongation:

1. **[prov]** prolongation of a vowel, nasal, lateral, or approximant, e.g. [j:es] y**[prov]**es; [n:əʊ] n**[prov]**o. The 'v' in [prov] refers to vowels since these are the most commonly occurring prolonged segment.

2. **[prof]** prolongation of a fricative, e.g. [s:ɪks] s**[prof]**ix

In 2 above only one syllable is counted regardless of the duration of the prolongation. If it were [s:.sɪks] it would be transcribed as s**[prof]**- six**[pwr]**). As the coding indicates, this is a combination of two types of disfluency, and two syllables would be counted.

3. **[prop]** prolongation of a plosive. This applies when:

3a) The closed phase of a plosive lasts  $\geq 200$  msec, e.g. [k:a:] c**[prop]**ar. This does not include aspiration. There may be some attendant vocalisation during a prolonged plosive, but this is not as critical as the duration. For a plosive in utterance-initial position, however, the duration of closure may be difficult to assess and, in such cases, attendant vocalisation along with  $\geq 200$  msec pause before plosive release suggests that the speaker is failing to release a plosive rather than deliberately delaying the onset of an utterance.

3b) A word-final plosive is released relatively slowly and thereby becomes more fricative-like. However, as the origin of the sound is a plosive these are also classified as **[prop]**, e.g. [bakx:] *back***[prop]**. This involves aspiration because the plosive is being released but the release phase is considerably longer than might be expected.

3c) An affricate exhibits an extended closed phase or an unusually long release, e.g. [tɔ:t:f] *torc* **[prop]**h and [tɔ:t:f:] *torch***[prop]**.

**2.5.4 Note:** If more than one prolonged segment occurs in a single word each prolongation is coded separately, e.g. [tɔ::t:f:] *tor***[prov]***ch***[prop]**

## 2.6 Interruptions

**2.6.1 Definition:** The speaker interrupts him/herself and changes what had been started. A correction may be made.

**Note:** alternative taxonomies may code interruptions as deletions, insertions or repairs. However, given the limited number of interruptions arising in the present data, the use of a

finer grained analysis in which the interruption was ‘explained’ rather than described would be unwieldy.

**2.6.2 Types:** There are two main types of interruption and a variant:

1. **[pint]** interrupted phrase, e.g. *I was going to come at- [pint] I was going to leave early.* Repetitions following an interruption are not coded if a word or words at the start of the phrase prior to the INT are repeated in the phrase following the INT, but the later word or words of these two phrases differ as in the above example.

2. **[wint]** interrupted word, e.g. *ye-[wint] no*, or a response to a factual query such as *What was she called?* Answer: *Jea-[wint] Joan.* The incomplete words are counted as complete syllables.

3. **[wint][pint]** variant: interrupted word and phrase, e.g. *I was going to k-[wint][pint] I decided to leave at ten.* A word has been started then interrupted so that both a word and the phrase containing the word were interrupted. The [k] in this example is counted as a whole syllable. There may be more than one syllable realised, e.g. *I expe-[wint][pint] I decided not to go.* In this case the two syllables of the incomplete word will be counted.

### 3 METHOD

#### 3.1 Speakers and recordings

The speakers were 20 male speakers of Standard Southern British English (SSBE), aged 18-25 from the *DyViS* database (Nolan, McDougall, de Jong, & Hudson, 2009), recorded between March 2006 and August 2007 at the University of Cambridge. The speakers had no history of speech or hearing problems and their status as speakers of SSBE was judged by a phonetician who is a native speaker of that variety. The speakers are numbered S1, S2, S3, ..., S23 (no S5, S7, S14). The *DyViS* database provides recordings of 100 male SSBE speakers undertaking tasks in several speaking styles. The present study uses recordings of Task 1, a simulated police interview, an extension of the map task technique (Anderson *et al.*, 1991) in which the experimenter assumed the role of police officer and the speaker was the suspect being questioned. The interviewer and speaker were facing computer screens displaying a *PowerPoint* presentation controlled by the experimenter. The speaker was instructed that his memory and knowledge were represented in the maps and schemas that would appear on the screen and that he should answer the police officer’s questions using the information shown in black on the screen, but that he should avoid mentioning or deny the incriminating information shown in red. He was told that it was also okay to say ‘I don’t know’ or ‘I can’t remember’ when being questioned about red information. The interviewer’s questions guided the speaker through a crime scenario shown on the slides, asking him to describe where he was and what he did on the day of the crime and questions about his home, workplace, colleagues, friends, etc. Similarly to the map task, the slides were designed to elicit a number of target phonetic variables, but these are not of concern to the present study. The methodology introduced an element of cognitive challenge into the task in the use of material which could be freely spoken about and topics which the speaker was asked to conceal. The questions asked were essentially the same for each participant though there were variations in wording and timing which naturally occur in speech. Further details about the *DyViS* database and its elicitation techniques are available in Nolan *et al.* (2009).

### **3.2 Measuring the fluency phenomena**

The metric by which each speaker's performance was compared with those of other speakers was the number of occurrences each type of disfluency per 100 syllables of speech. All syllables were detected and counted manually. A strictly phonetic approach is used in that all syllables are counted even if the word/words are repeated or if the syllable is incomplete, e.g. in a part-word repetition (see below). Prolonged syllables are counted as single syllables. Filled pauses and back channel utterances such as 'mmm' are not included in the syllable count. However, single words such as 'yes' or 'no' are included. The data set examined consists of ten 100-syllable sets for all speakers except S15, S16, S17 and S22 for whom nine 100-syllable sets are available and S13 for whom there are eight 100-syllable sets.

### **3.3 Transcribing the phenomena**

*Praat* (Boersma & Weenink, 1992-2016) was used to examine the audio recordings of the interviews. Text grids were used to transcribe the speech orthographically and to code the fluency phenomena using the taxonomy described in Section 2 above.

### 3.4 Annotating and reviewing the transcriptions and tabulating the results

The transcriptions in the *Praat* text grids were transferred to an Excel spreadsheet. The boundaries of an utterance were often a filled pause or unfilled pause though longer utterances could be broken up at convenient points. The spreadsheet contained columns for the total number of syllables and for each of the 14 different types of phenomena:

Filled and unfilled pauses: **[er]**, **[erm]**, **[fpo]**, **[pg]**, **[po]**;

Repetitions: **[wrep]**, **[pwr]**, **[prep]**, **[mrep]**;

Prolongation: **[prov]**, **[prof]**, **[prop]**;

Interruptions: **[pint]**, **[wint]**

The combined **[pint][wint]** variant was scored by marking both **[pint]** and **[wint]** present in the respective columns of the spreadsheet.

The spreadsheet also included an **[all]** column in which the sum of all separate fluency features was entered.

The initial analysis was undertaken by the first author. Even with careful definitions of the disfluency types human judgement was being relied upon for the coding and the syllable counts. An inter-analyst consistency study was carried out as described in Section 4 below.

## 4 CONSISTENCY TESTING

Given that there is a degree of subjective judgement involved in identifying and categorising the disfluency types analysed, a study was undertaken to test the degree of inter-analyst consistency. A subset of 5 speakers was reanalysed by two additional analysts and the consistency of disfluency feature measurements across analysts was evaluated.

The two new analysts (the first author, and a speech and language therapist) undertook training with the first analyst (the second author) to become familiar with the criteria for identifying each disfluency type, using the coding system, and the syllable counting protocol. At a subsequent meeting, the analysts discussed their experiences of using the categorisation system and jointly decided on revised criteria (as presented in the previous section) for the identification of features which had proved ambiguous or problematic. Each analyst then worked independently, reviewing his or her own coding record using the revised categorisation criteria. The total number of counts recorded by each analyst for each feature, combining the ten 100-syllable sets for the five speakers, is shown in Table 1.

*Table 1. Total number of counts recorded by the three analysts for each disfluency type, combining the ten 100-syllable sets for the five speakers.*

Analyst	er	erm	fpo	pg	po	pwr	wrep	prep	mrep	prov	prof	prop	pint	wint	all
1	252	261	2	353	144	30	125	18	11	94	27	32	51	22	1422
2	261	252	4	374	80	21	105	16	10	19	1	1	34	32	1210
3	257	241	15	438	77	27	116	18	9	64	14	9	31	19	1335

The counts per 100-syllable set for each disfluency type were subjected to Pearson's correlation analysis, comparing each pairing among the three analysts. Results are given in Table 2 below.

Table 2. Pearson's correlation coefficients among the disfluency counts made by Analysts 1 and 2, 1 and 3, and 2 and 3.

Pair of Analysts Tested	er	erm	fpo	pg	po	pwr	wrep	prep	mrep	prov	prof	prop	pint	wint	all
1~2	0.87	0.88	0.05	0.80	0.59	0.74	0.84	0.75	0.81	0.35	0.08	0.53	0.61	0.65	0.88
1~3	0.88	0.91	0.09	0.86	0.47	0.86	0.88	0.74	0.74	0.63	0.41	0.52	0.63	0.57	0.93
2~3	0.86	0.84	0.25	0.88	0.47	0.80	0.87	0.83	0.72	0.30	0.19	0.50	0.59	0.62	0.89

#### 4.1 Filled pauses: [er], [erm], [fpo]

[er] and [erm] yielded high levels of correlation among all three pairs of analysts (ranging between  $r = 0.84$  and  $0.91$ ), indicating a good degree of inter-analyst agreement in the identification of these categories. Correlations among [fpo] counts were on the contrary very low among the three analysts. However, the total number of [fpo] features identified was also very low (2-15 for the three analysts; compared with 252-261 for [er] and 241-261 for [erm]). Although this suggests that [fpo] category is may be less helpful in comparing speakers the fact that it is relatively infrequent and possibly idiosyncratic could nevertheless mean that it has probative value.

#### 4.2 Silent pauses: [pg], [po]

The silent pause category [pg] (pause at a grammatical boundary) yielded high levels of correlation among all three pairs of analysts (ranging between  $r = 0.80$  and  $0.86$ ), with 353-438 [pg] pauses having been identified. The 'other' category of pause, [po], gave more moderate levels of correlation (ranging between  $r = 0.47$  and  $0.59$ ), indicating a lower extent of correspondence among the three analysts in identifying pauses which were not at grammatical boundaries. 77-144 [po] pauses were identified by the analysts. When the two pause categories are merged and the counts combined for [pg] + [po], high levels of correlation among the analysts are achieved ( $r = 0.80$  to  $0.92$ ), suggesting that it is the type of pause categories about which the analysts do not always agree, rather than the identification of a pause itself. In 2.2.2 it was noted there may be speech phenomena which militate against a simple definition of whether a pause is [pg] or [po] and therefore the distinction may not be sufficiently robust for the purposes of TOFFA. This will be taken into account in future research.

#### 4.3 Repetitions: [pwr], [wrep], [prep] and [mrep]

High levels of correlation were exhibited among the three pairs of analysts in their counts for all four of the repetition categories (*pwr*:  $r = 0.74$  to  $0.86$ , *wrep*:  $r = 0.84$  to  $0.87$ , *prep*:  $r = 0.74$  to  $0.83$ , *mrep*:  $r = 0.72$  to  $0.81$ ), indicating a good level of agreement among the analysts in the identification of repetition types.

#### 4.4 Prolongations: [prov], [prof] and [prop]

Levels of correlation among the various types of prolongation were lower than all other disfluency types, apart from [fpo] ([prov]:  $r = 0.30$  to  $0.63$ , [prof]:  $r = 0.08$  to  $0.41$ , [prop]:  $r = 0.50$  to  $0.53$ ). From the totals shown in Table 1, it is clear that Analyst 1 identified the

occurrence of prolongations far more frequently than Analysts 2 and 3. Analysts had been required to include articulations  $\geq 200$  msec. This includes prolongations motivated by prosody such as utterance utterance-final lengthening. These may be relatively harder to identify because they are functionally appropriate. On the other hand, prolonged aspects of plosives [**prop**] were relatively highly correlated across the analysts. These occurred relatively less often but their pause-like quality may have made them more salient. Further, although the number of counts seen in Table 1 are quite different between Analysts 1 and 2 and between 1 and 3, fairly high correlations are yielded due to the large number of 100-syllable groups identified as containing no [**prop**]. The lowest correlation of all relates to [**prof**] but this was not consistent across the comparisons between the analysts. Eklund (2004) noted that prolongations formed the third most common disfluency category in his study of Swedish and that prolongations occurred in the NF speech of speakers in many languages. Because these phenomena do not necessarily disrupt the overall rhythm of speech and may indeed contribute to the meaning being conveyed they may be relatively harder to identify consistently.

The co-analysts in the present study commented upon the relative difficulty of identifying prolongations. It may be cognitively demanding to identify a feature which both affects the overall flow of an utterance and may also convey meaning. The decision to use  $\geq 200$  msec as the criterion for identifying a prolongation may also be problematic. In the note in 2.5.2. a number of different approaches to identifying prolongations are made. These will be taken into consideration in future research.

#### **4.5 Interruptions [**pint**], [**wint**]**

The interruption categories showed moderately high levels of correlation among the three analysts ([**pint**]:  $r = 0.59$  to  $0.63$ , [**wint**]:  $r = 0.57$  to  $0.65$ ).

#### **4.6 All fluency features combined: [**all**]**

This yielded high levels of correlation among all three pairs of analysts at  $r = 0.88$  to  $0.93$ , confirming that fluency feature analysis if all the features are considered exhibits a reasonable level of inter-analyst consistency.

#### **4.7 Inter-analyst consistency studies in other fluency feature research**

Eklund (2004), Braun and Rosin (2015) and Schiel and Heinrich (2015) present no inter-analyst consistency data. Roberts *et al.* (2009) conducted a consistency study between two analysts. They achieved high inter-rater agreement though they only calculated it for the overall number of fluency features, equivalent to [**all**] in the present study, which also shows a relatively high level of correlation among analysts. Roberts *et al.* did not provide numerical consistency measures for the sub-types of fluency features though they comment that “agreement on the frequency sub-types was lower, but still judged acceptable” (2009: 418). The total number of occurrences of these sub-features are however considerably less than in the present study.

## **5 RESULTS**

In this section the results used are those from Analyst 1, the second author.

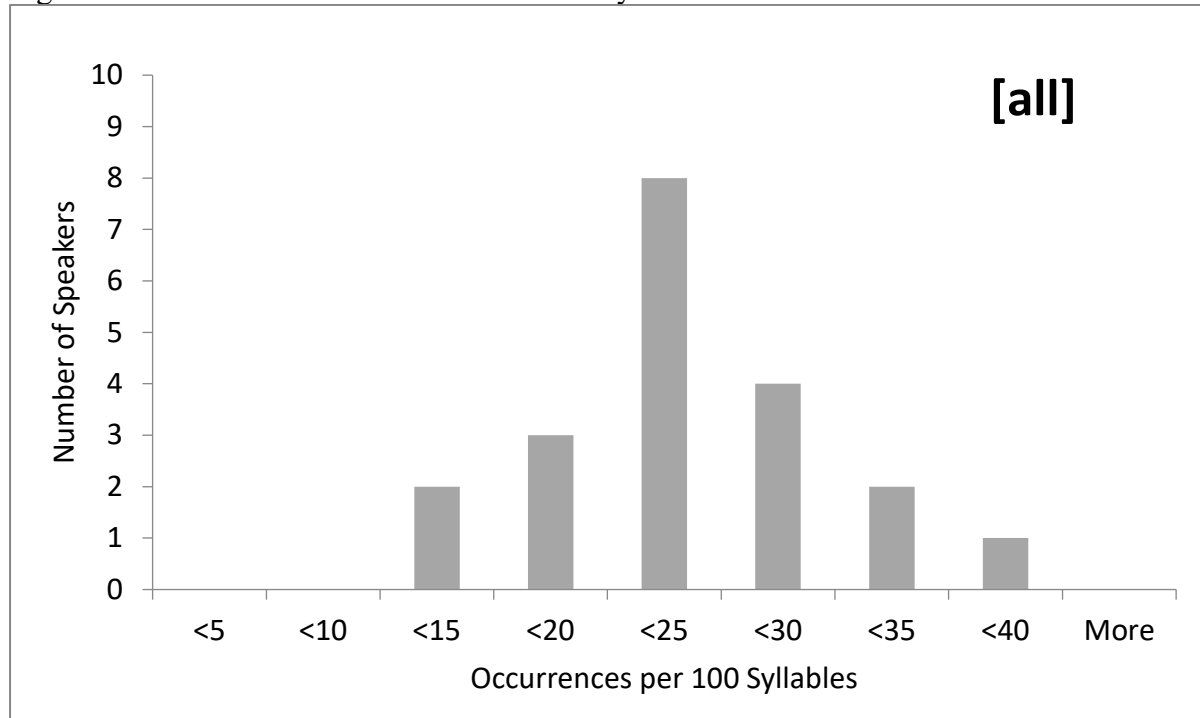
### **5.1 Graphical analysis**

The numerical data was presented in a series of histograms showing the number of speakers who produced each type of fluency feature per 100 syllables

### 5.1.1 Overall rates of fluency features [all]

The distribution of the 20 individuals' overall rates of production of fluency features are shown in Figure 1.

Figure 1. Individuals' overall rates of all fluency features.



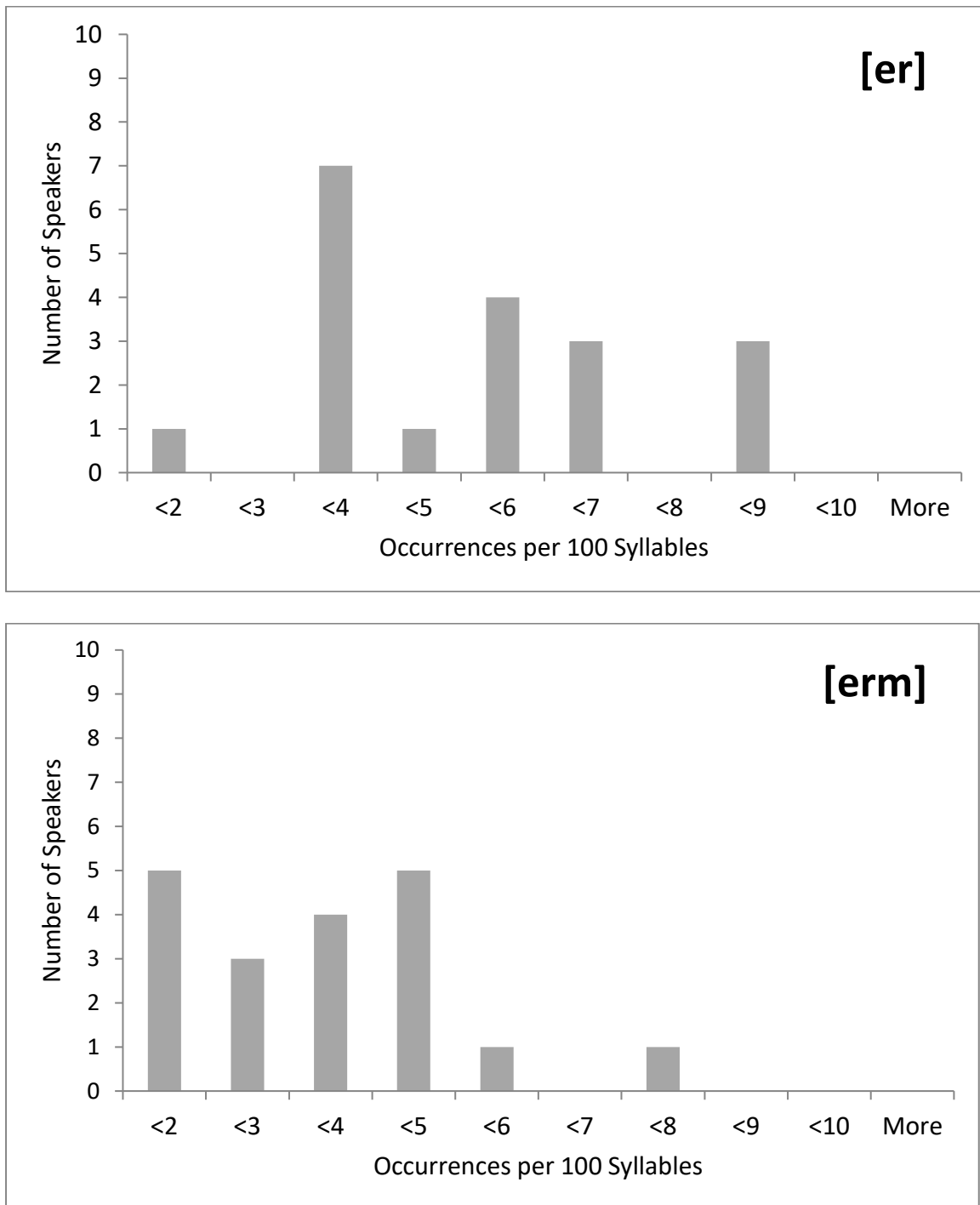
A wide range of individual variation is present, with rates ranging from 13.0 to 36.8 fluency features per 100 syllables. The data form an approximately normal-shaped distribution, with the majority of speakers occurring within the middle two bins (<25 and <30 occurrences per 100 syllables) and fewer speakers' rates appearing in the less central bins. Histograms showing the speakers' rates of occurrence for each separate type of disfluency also display individual variation, with certain features appearing to exhibit greater individual differences than others, as is discussed in the following sections.

### 5.2.2 Filled pauses

Speakers' rates of production of the filled pauses [er] and [erm] are shown in Figure 2. The occurrence of other filled pauses [fpo] was very infrequent, with most speakers not using them at all, thus the results for this feature are not discussed further here. This does not however indicate that they are of little forensic significance as indicated in 4.1.

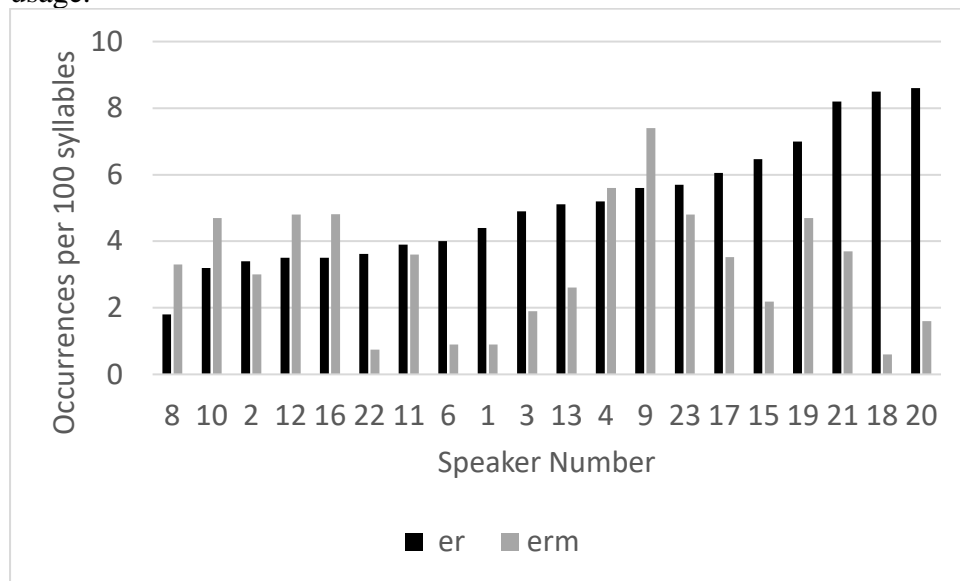


Figure 2. Individuals' rates of occurrence of the filled pause **[er]** (upper panel) and **[erm]** (lower panel).



The filled pause **[er]** occurs on the whole more often than **[erm]**, but with individuals' rates of usage varying widely for both types. Usage rates of these two types of pauses vary considerably between individuals, with some speakers employing more **[er]** than **[erm]**, others more **[erm]** than **[er]**, and others again using similar rates of both, as shown in Figure 3. The lack of consistent pattern in individuals' use of **[er]** and **[erm]** is confirmed by the low correlation between the two rates ( $r = -0.12$ ).

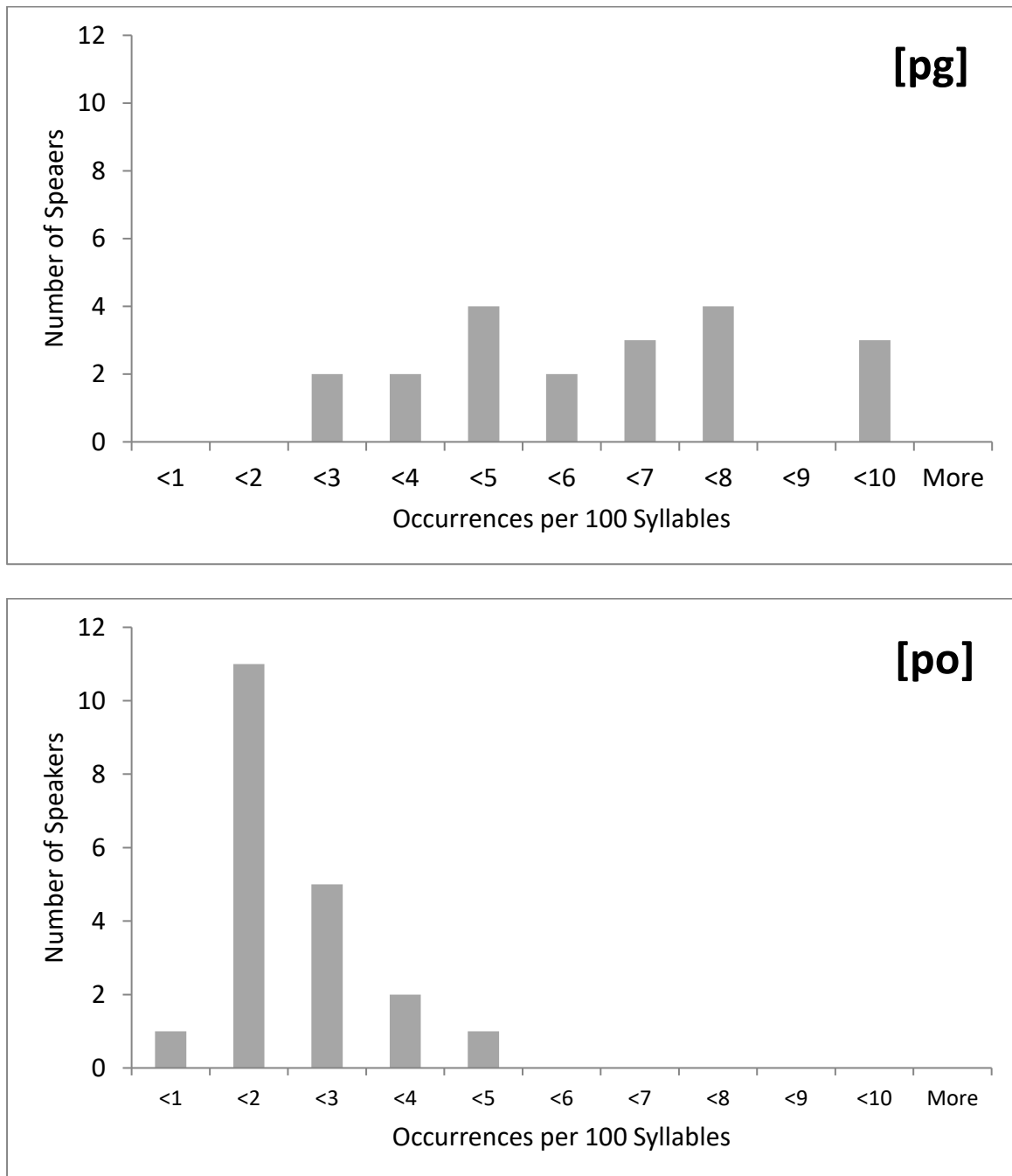
Figure 3. Individuals' mean rates of usage of [er] and [erm] per 100 syllables, ranked by [er] usage.



### 5.3.3 Silent pauses

Speakers' rates of silent pauses are shown in Figure 4. Rates of occurrences of [pg] are relatively evenly distributed across the bins between <3 and <10 occurrences per 100 syllables. For [po], most speakers' rates occur in the <2 or <3 bins. This reveals how rarely [po] events occur compared to [pg] events. It is possible that the range of rates demonstrated for [pg] reflects differences between speakers, however, it may be that the granularity of the bin-sizes used is too small. For both variables, more data is needed to establish whether the apparent distribution curves are robust.

Figure 4. Individuals' rates of occurrence of grammatical silent pauses [pg] (upper panel) and other silent pauses [po] (lower panel).



#### 5.4.4 Repetitions

Speakers' rates of occurrence of the three types of repetition as well as multiple repetitions are shown in Figure 5. Part-word repetitions [pwr] were produced rarely by most speakers, with 17 of the group of 20 producing 0 or <0.5 part-word repetitions per 100 syllables. Three speakers, however, produced between 0.5 and 1.5 part-word repetitions per 100 syllables. These are also low rates of occurrence, but might indicate that the usage of this type of disfluency is a speaker-specific feature of the speech of these speakers. In a forensic speech analysis therefore, an occurrence of more than two part-word repetitions per 100 syllables in a disputed speech sample would be relatively distinctive.

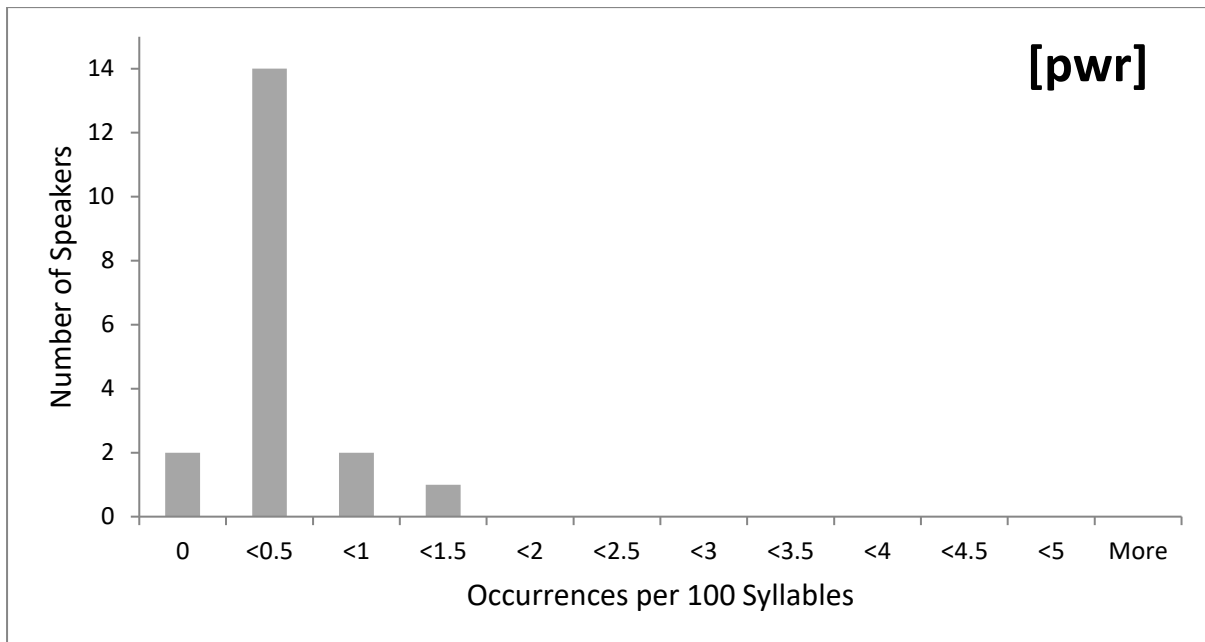
Repetition of whole words [**wrep**] offers a greater range of individual differences among the group of speakers, with seven speakers producing less than 0.5 per 100 syllables, and the remaining 13 speakers producing a relatively even spread of results across the 0.5 bins up to <4.5.

Phrase repetitions [**prep**] display a pattern of usage similar to part-word repetitions, with the majority of speakers (15) producing 0 or <0.5 [**prep**] per 100 syllables, and a small number of speakers (5) producing between 0.5 and 2 [**prep**] per 100 syllables.

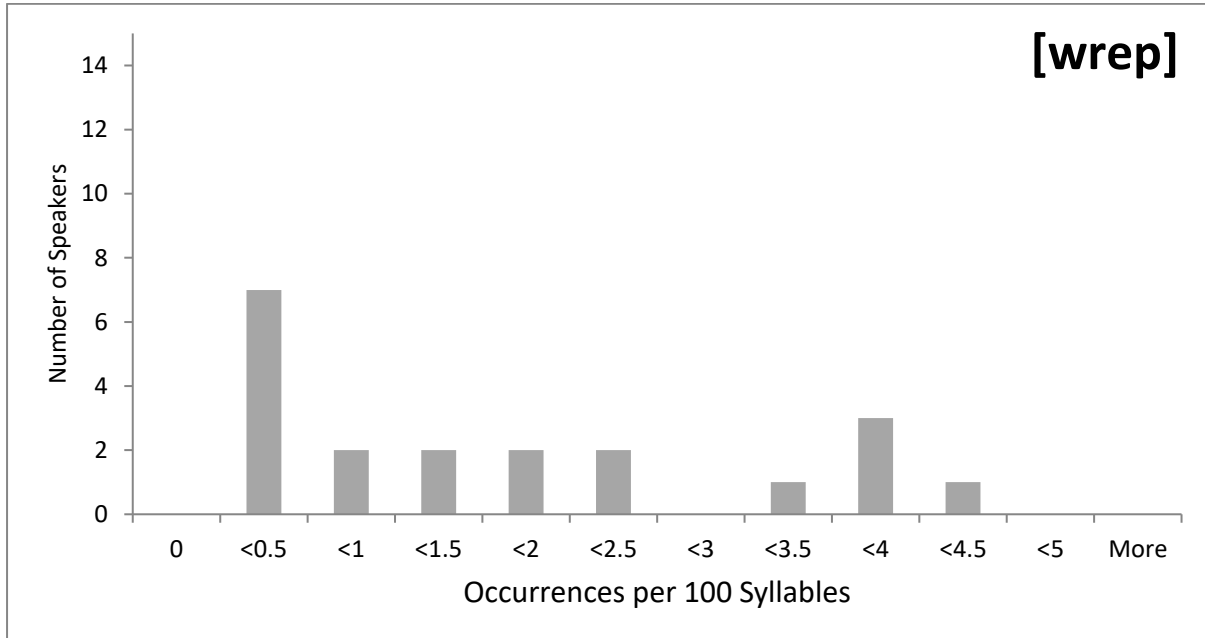
Multiple repetitions [**mrep**] occurred infrequently. From a forensic perspective, the occurrence of one or two instances of [**mrep**] on a disputed recording would not offer much in terms of speaker-specificity, because although this is a less commonly occurring disfluency in the NF speakers, it is by no means an atypical type of fluency feature. An occurrence rate of >1 [**mrep**] per 100 syllables would however be distinctive.

Figure 5. Individuals' rates of occurrence of (i) part-word repetitions, (ii) word repetitions, (iii) phrase repetitions, (iv) multiple repetitions.

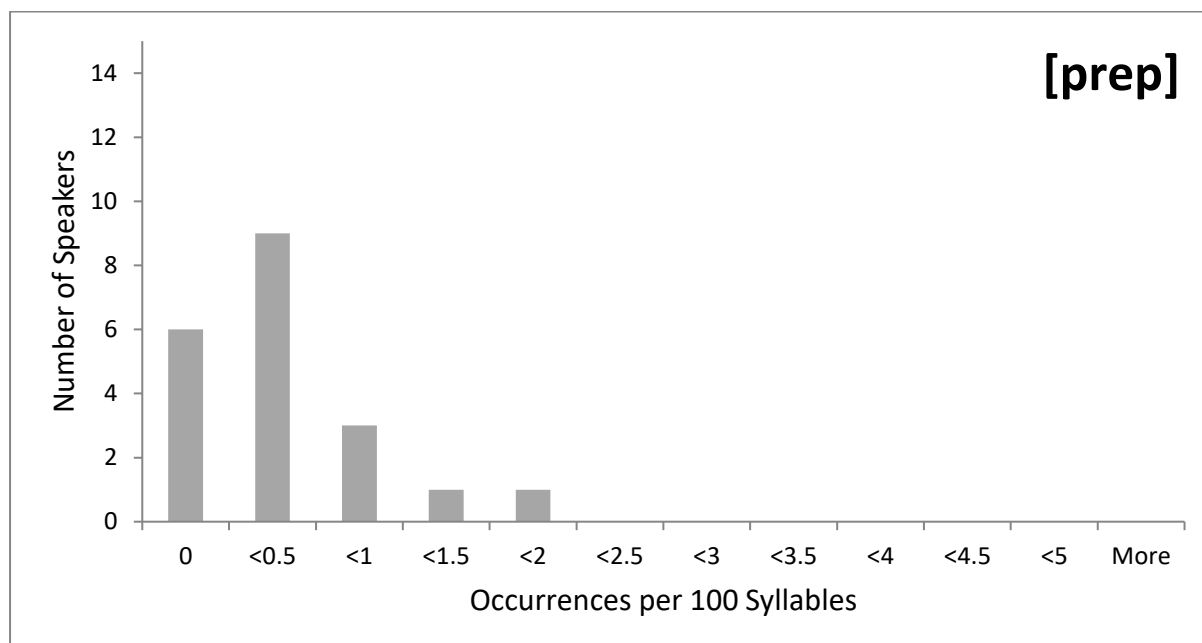
**(i) part-word repetitions**



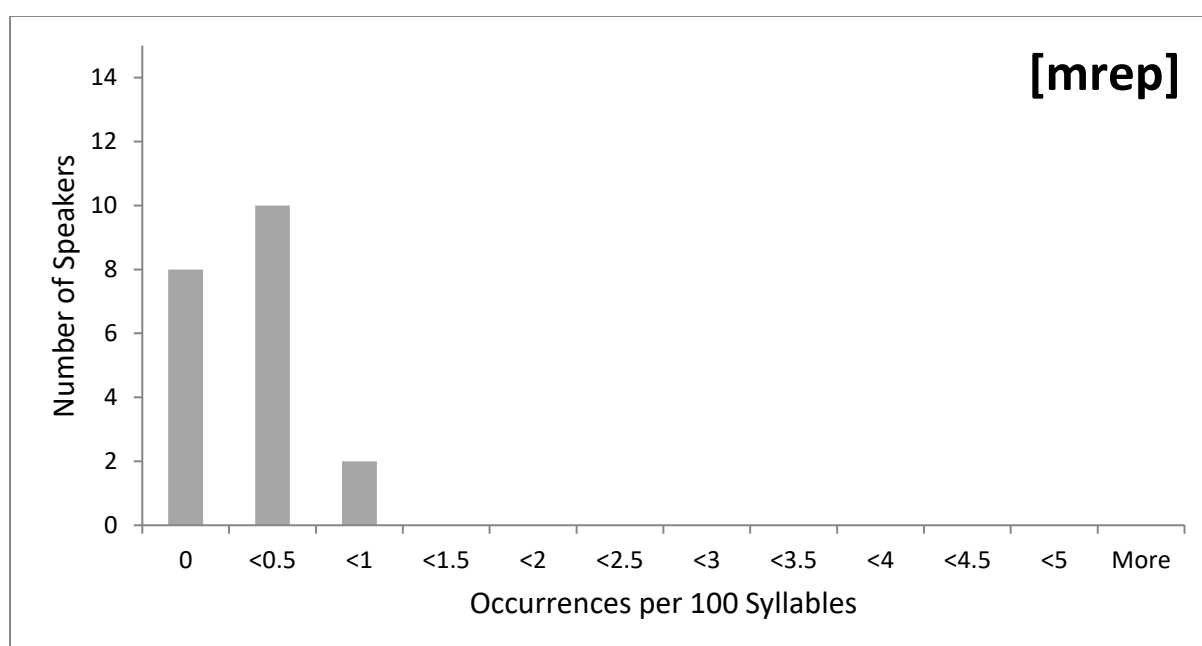
**(ii) whole word repetitions**



**(iii) phrase repetitions**



**(iv) multiple repetitions**



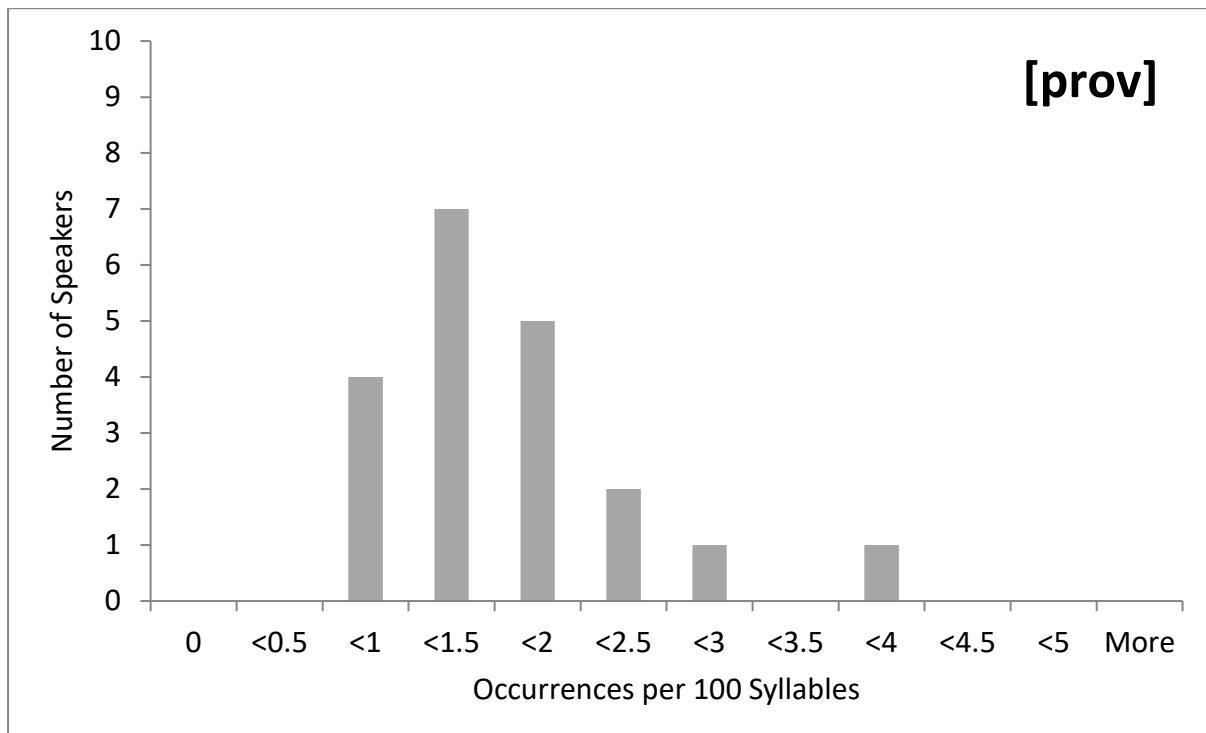
**5.4.5 Prolongations**

Speakers' rates of occurrence of the three types of prolongation are shown in the histograms in Figure 6. For all three types, the data appear to form a positively-skewed distribution, with most speakers' rates falling in the <0.5 and <1 bins (**[prof]** and **[prop]**) or <1 and <1.5 bins (**[prov]**), and decreasing numbers of speakers' rates occurring in higher bins. Prolongations in themselves are a regular part of NF speech not least because many of the prolongations

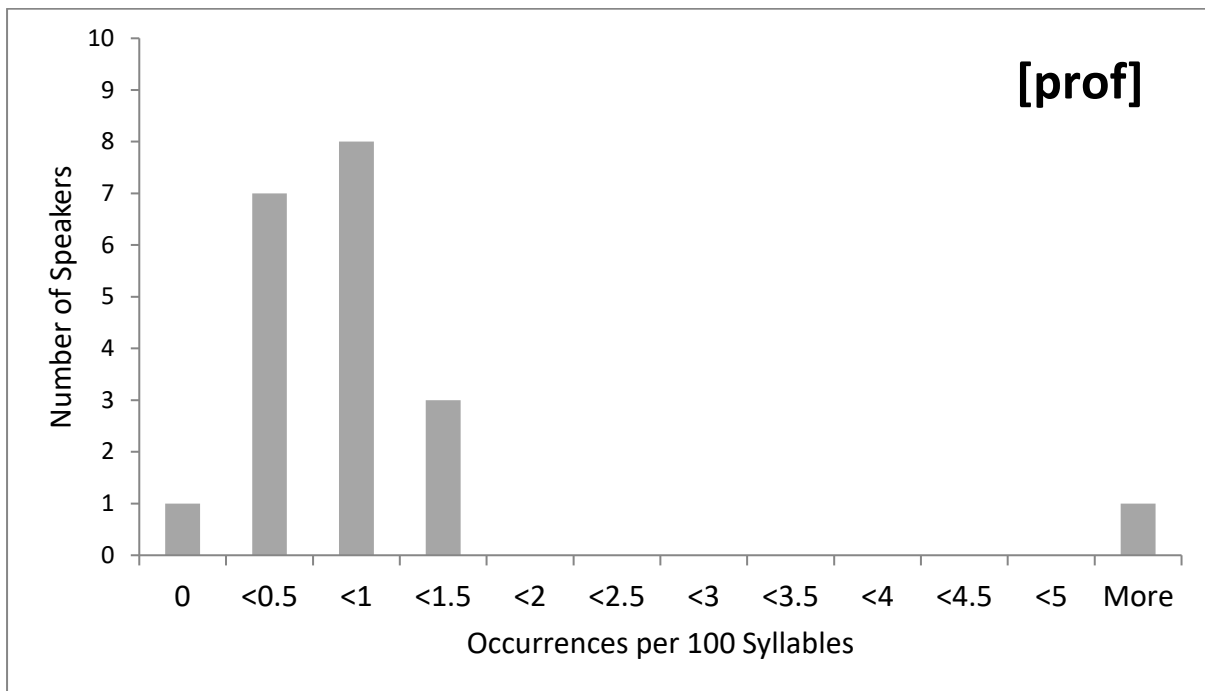
have a specific linguistic function. In the forensic domain, >3 **[prov]** occurrences per 100 syllables, or >2 **[prof]** or **[prop]** occurrences per 100 syllables would be distinctive because very few of the 20 speakers within the present sample use these features more frequently than this.

Figure 6. Individuals' rates of occurrence of (i) vocalic prolongations, (ii) fricative prolongations, and (iii) plosive closure/affricate closure or release prolongations.

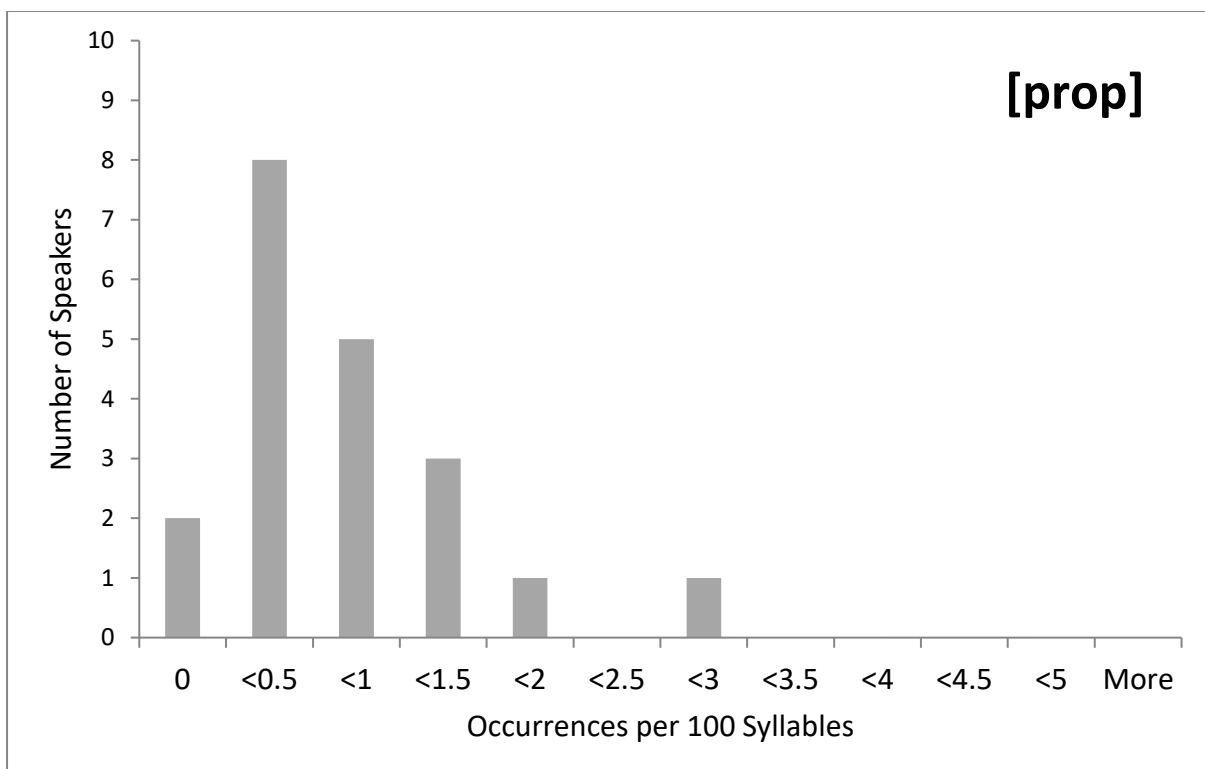
**(i) vocalic prolongations**



**(ii) fricative prolongations**



**(iii) plosive/affricate closure or release prolongations**



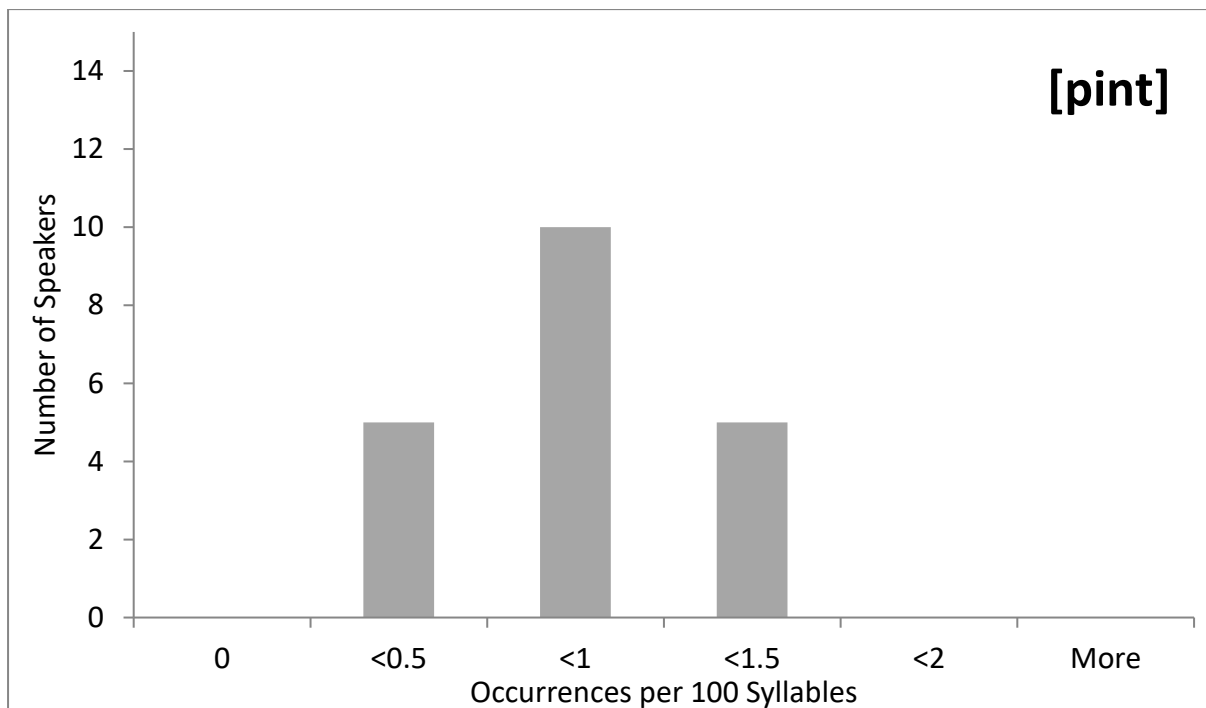


#### 5.4.6 Interruptions

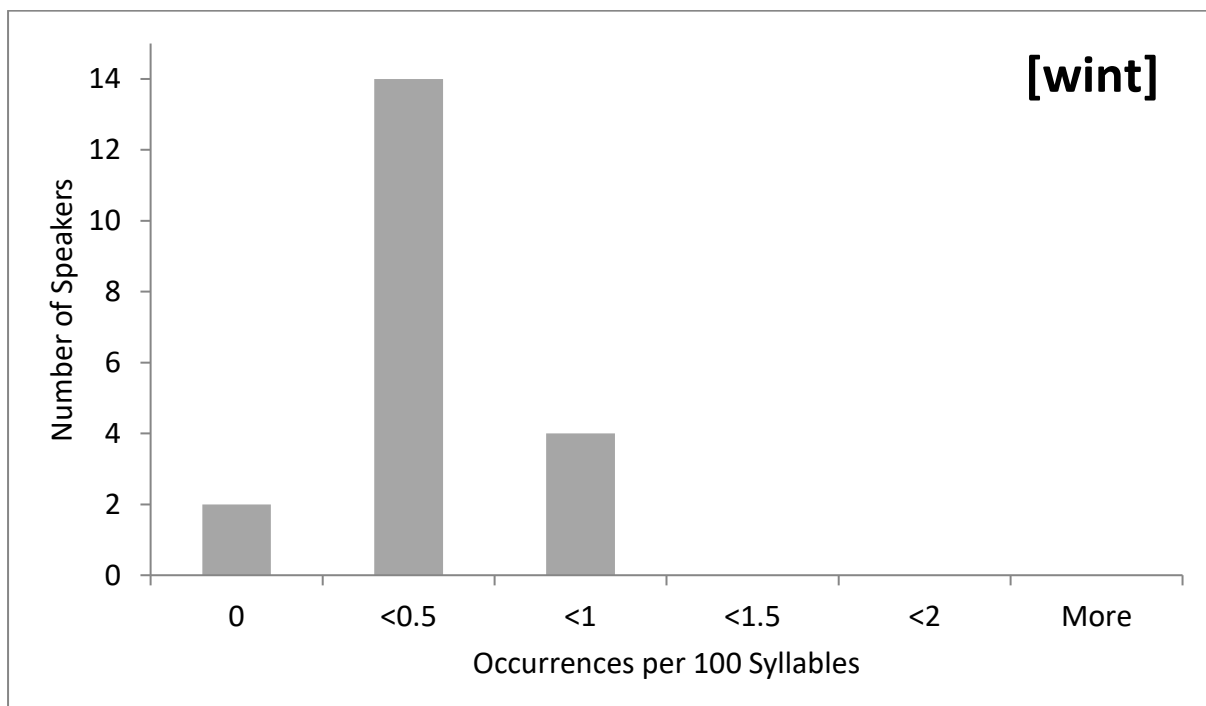
Speakers' rates of occurrence of the two interruption types are shown in Figure 7. All speakers produced both phrase and word interruptions, but their occurrence in the present sample never exceeded two for phrase interruptions [**pint**] or one for word interruptions [**wint**] per 100 syllables. Interruptions do not appear to offer a wide range of individual differences, though occurrence rates of  $\geq 2$  per 100 syllables are potentially distinctive.

Figure 7. Individuals' rates of occurrence of (i) phrase interruptions, and (ii) word interruptions.

**(i) phrase interruptions**



**(ii) word interruptions**



## 6 PROFILING DISFLUENCIES

Overall mean rates for each of the disfluencies used by the 20 speakers are shown in Figure 8. Filled and unfilled pauses occur most frequently in the sample with repetitions, prolongations and interruptions occurring commonly, but less frequently. Extensive between-speaker variation in fluency behaviour is present in the disfluency profile used by each individual and the extent to which each feature is used. As an example, Figure 9 presents the disfluency profiles for two individuals whose overall rate of disfluency occurrence is very similar. Although both speakers produce approximately 23 disfluencies per 100 syllables, their profiles are very different from each other. Speaker 12 uses more **[erm]** than **[er]**, while Speaker 13 does the reverse, and the overall numbers of filled pauses used also differ between the speakers. Speaker 13 uses more silent pauses than Speaker 12 (9.1 vs 7.8 per 100 syllables). Both speakers use more **[pg]** than **[po]**, but Speaker 12 uses more **[pg]** (6.8) than Speaker 13 (5.3). Speaker 13 exhibits little use of repetition, while Speaker 12 produces **[wrep]** at a rate of 2.0 per 100 syllables. The two speakers employ different patterns of prolongations, with Speaker 13 showing higher levels of **[prov]** usage (2.5) than **[prof]** (1.1) or **[prop]** (0.8) and Speaker 12 showing more even use of the three types (0.7-1.3). Both speakers use interruptions infrequently, consistent with the overall behaviour of the group described earlier.

Figure 8. Mean rates of occurrence of each fluency feature per 100 syllables, across the 20 speakers. Shading indicates membership of the categories Filled Pause (light grey), Silent Pause (backwards diagonal), Repetition (dark grey), Prolongation (horizontal stripes) and Interruption (forwards diagonal).

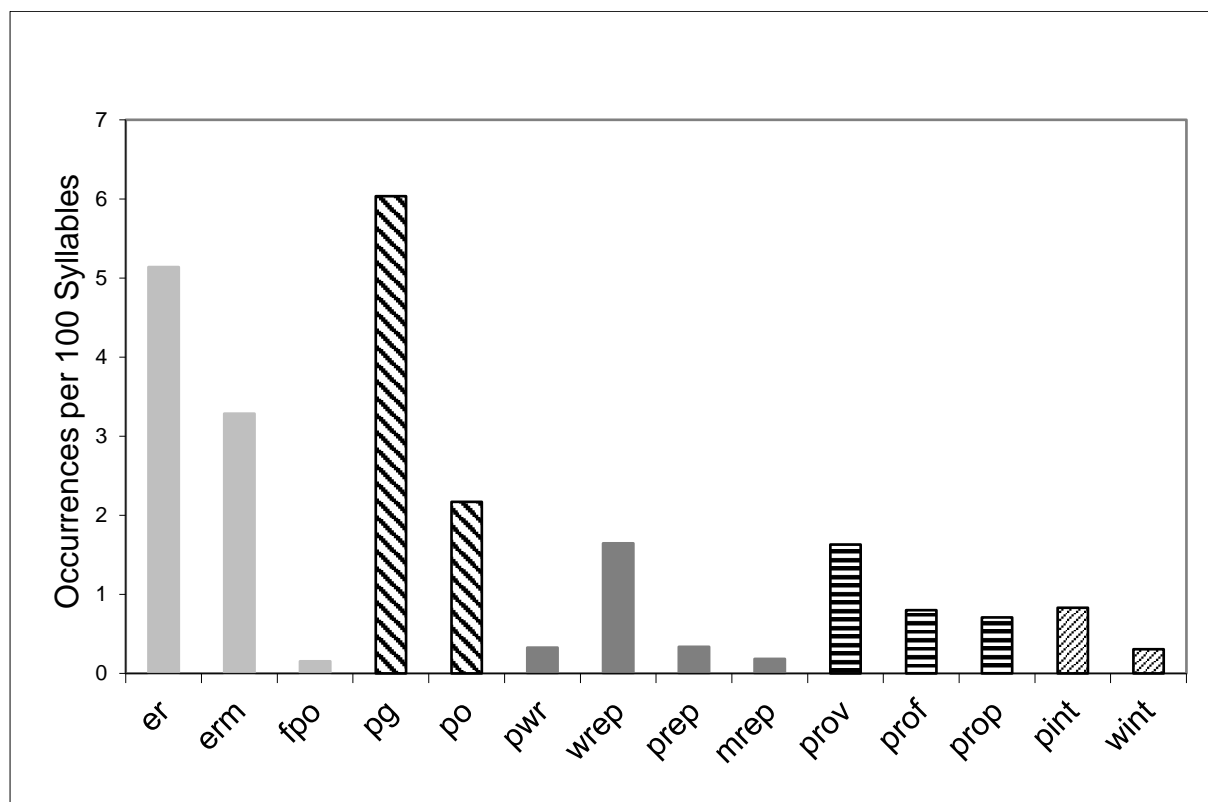
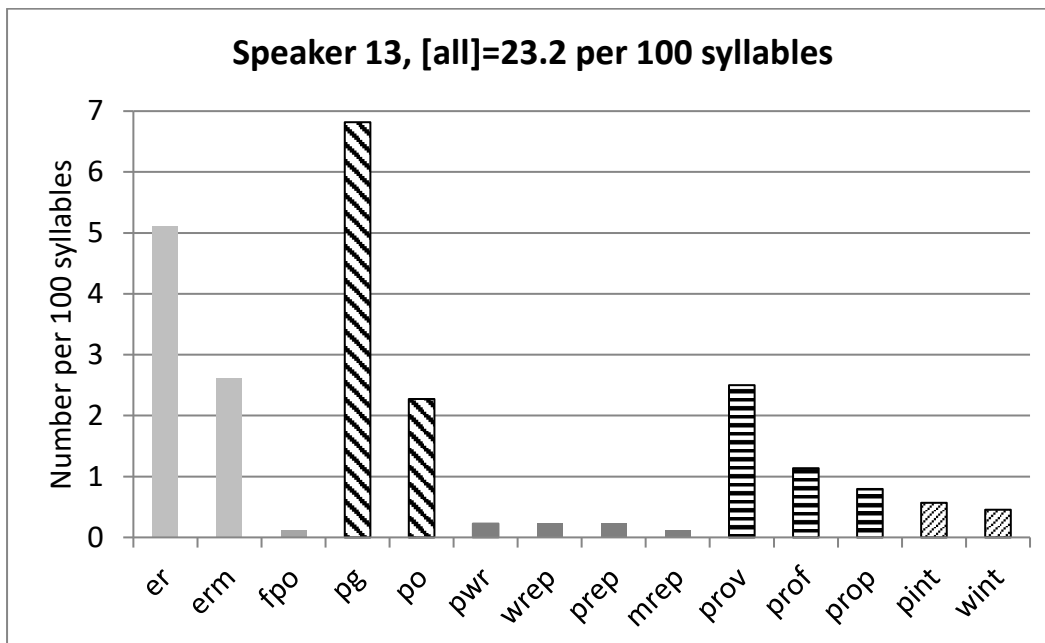
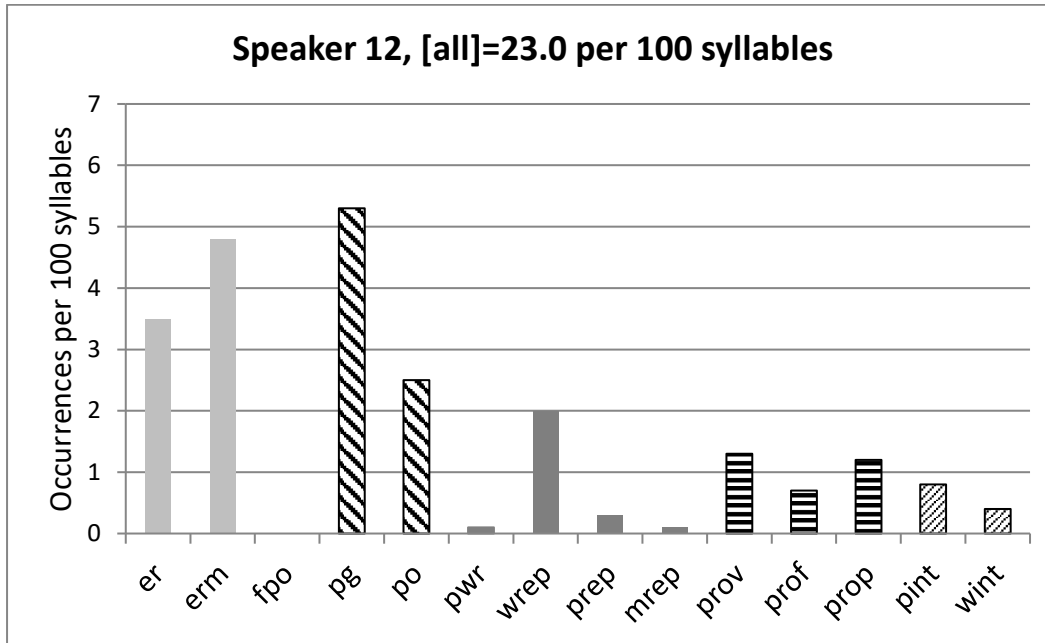


Figure 9. Comparison of fluency profiles for two speakers with very similar overall fluency feature rates, Speaker 12 (upper panel; [all] = 23.0) and Speaker 13 (lower panel; [all] = 23.2). Shading indicates membership of the categories Filled Pause (light grey), Silent Pause (forward diagonal), Repetition (dark grey), Prolongation (horizontal stripes) and Interruption (backwards diagonal).



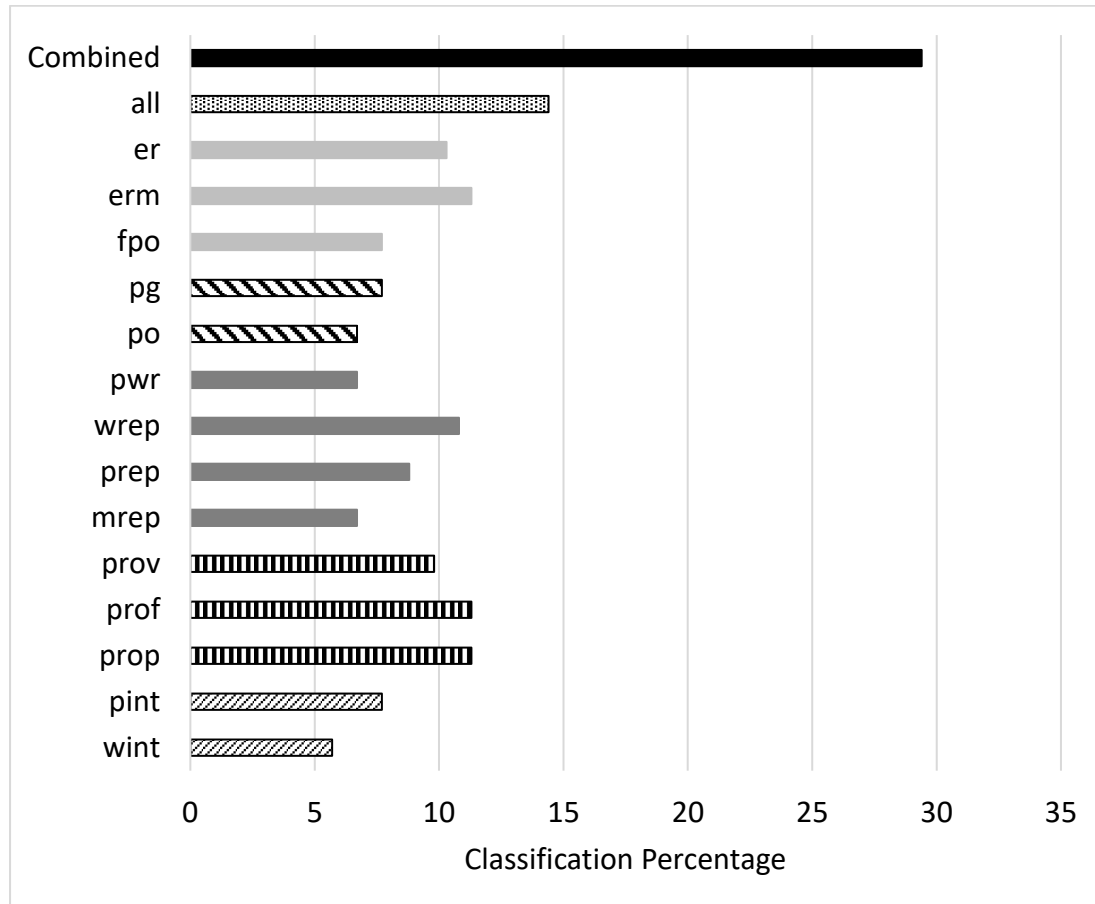
## 7 DISCRIMINANT ANALYSIS

To test statistically the extent of speaker-specificity exhibited by each disfluency type and by disfluencies in combination, discriminant analysis was used for the 20 speakers under examination. This analysis is a multivariate technique which can be used to determine whether a set of predictors can be combined to predict group membership (Tabachnick & Fidell, 2014: ch. 9). It is acknowledged that a study with a focus on the analysis of a speech variable purely for application in forensic casework should involve a likelihood ratio approach. However, the present study is concerned with individual differences in speech production more broadly than for forensic application only. Further, while it may be possible to carry out a likelihood ratio analysis with just 20 speakers, the results would be more effective with a larger population of speakers (see e.g. Hughes, 2014; Ishihara & Kinoshita, 2008). The discriminant analysis presented here provides an initial statistical exploration which can be built on to incorporate likelihood ratio analysis in future larger-scale studies.

When carrying out discriminant analysis for a speaker discrimination study involving continuous variables, (e.g. an investigation of vowel formant frequencies such as McDougall, 2004, 2006), a 'group' is a speaker, or rather the set of utterances or tokens produced by a speaker. In the case of disfluencies, where the predictors are rates of occurrence per 100 syllables, the present study defines a speaker 'group' as the collection of 100-syllable sets produced by that speaker. The discriminant analysis procedure constructs discriminant functions which can be used to allocate each 100-syllable set in the data to one of the speakers and determines a 'classification rate' according to the accuracy of the allocation. In the present study, this is done using the 'leave-one-out' method, where each case is classified by discriminant functions derived from all cases except for the case itself.

Direct discriminant function analyses were performed for each disfluency type individually, using the rates of occurrence per 100 syllables as predictors of membership of twenty groups, S1, S2, S3, ... etc. ( $k = 20$ ). The resulting classification rates are shown in Figure 10.

Figure 10. Discriminant analysis classification rates for each disfluency type individually, for the [all] feature which adds all fluency features together, and for a combined analysis ('Combined') using the seven features that yielded the highest classification rates individually. Shading indicates membership of the categories Filled Pause (light grey), Silent Pause (backwards diagonal stripes), Repetition (dark grey), Prolongation (vertical stripes), and Interruption (forwards diagonal stripes). Chance level is 5% as there are 20 speakers.



The discriminant analyses yielded rates of classification of 5.7-11.3%, rates higher than chance ( $1/20 = 5\%$ ), for all individual fluency features, indicating that all the features analysed exhibit some degree of speaker-specificity, albeit clearly not sufficient to distinguish all twenty speakers from one another single-handedly. The [all] feature which sums the full set of features for each speaker produced a classification rate of 14.4%. The best performing individual features were the filled pauses [er] (10.3%) and [erm] (11.3%), word repetitions (10.8%), fricative prolongations (11.3%) and plosive prolongations (11.3%).

Following the discriminant analyses using the disfluency features individually as predictors, a combined analysis was carried out, using the seven predictors which had achieved the highest classification rates individually: [er], [erm], [wrep], [prov], [prof] and [prop]. (Seven were chosen as this is the maximum possible for this data set where the speaker with the smallest data set provided eight 100-syllable sets; for discriminant analysis the number of predictors must not exceed the number of items in the smallest set minus one. See Tabachnick and Fidell (2014: 425).) This combined analysis produced a markedly higher classification rate of 29.4%,

demonstrating the importance of considering speakers' fluency profiles for characterising differences among speakers.

## 8 DISCUSSION

The results presented here show that NF speakers exhibit a considerable range of disfluency behaviour. All speakers may hesitate, repeat, prolong and interrupt their speech in order to, for example, formulate and execute utterances, and to signal an intention in an interaction. Furthermore, such disfluencies assist listeners to comprehend what has been said and signal opportunities for them to respond. Disfluencies are natural phenomena which are useful both for the speaker to assist in planning and articulating speech, and for the listener in providing time for decoding the speaker's intentions.

The data examined here show that individual speakers of SSBE exhibit different overall rates of disfluency. Table 2 shows the data comparing previous studies presented earlier in Table 1, along with the finding from the present study.

Table 2. Overall disfluency rates produced by speakers (mean across the speakers, lowest rate for an individual, highest rate for an individual) in the present study and in previous studies. As in Table 1 the rates are in occurrence per 100 syllables except for Braun and Rosin (2015) which are per minute.

<b>Study</b>	<b>Language and speech task</b>	<b>Mean rate</b>	<b>Lowest rate</b>	<b>Highest rate</b>
Present Study	Simulated police interviews	10.9	6.2	17.7
Shriberg (2001)	English telephone conversations	6	1	18
Eklund (2004)	Swedish telephone conversations	6.4	1.6	16.1
Roberts <i>et al.</i> (2009)	English monologues	5.1	1	10
Braun and Rosin (2015)	German monologues	5.2 per min	4.5 per min	12.3 per min

The results in Table 2 suggest that the data from the mock police interviews in the present study yielded somewhat more disfluencies than in other studies. However, the comparisons are of course only partial because the studies listed used different taxonomies for identifying disfluency and different means of eliciting data. The pattern for the overall disfluency rates of individuals being spread across a range occurs for the present study, consistent with the previous ones. The data collected in the present study have enabled a very detailed examination of this individual variation through an analysis of individual differences in the production of the separate types of disfluency feature contributing to these overall rates, and an analysis of differences between the disfluency profiles of the individual speakers.

In the present study, the disfluency features yielding the greatest levels of speaker-specificity according to discriminant analysis are the filled pauses *er* and *erm*, prolongations, and word repetitions. Filled pauses display notable differences among speakers, with rates ranging from <2 to 8 (*er*) and <2 to 7 (*erm*) per 100 syllables. This is consistent with the findings of Braun and Rosin (2015) who used a visual profile to show that speakers with a similar total number of filled pauses may be distinguished from each other by the relative proportions of the types of filled pause they employ. Individual differences in choice of filled pause type are further demonstrated in the present study, for example in Figure 9, where the profiles of Speakers 12 and 13 reveal different preferences for *er* versus *erm*.

The histograms in Section 5 show that a spread of individual behaviours is clearly present for all three types of prolongation. Further, for fricative prolongations one speaker produces more than five occurrences per 100 syllables, while the remaining 19 speakers produce fewer than 1.5 occurrences per 100 syllables. The findings from the present study, together with the findings of Eklund (2004) for Swedish, and Braun and Rosin (2015) and Schiel and Heinrich (2015) for German, indicate that prolongations are one of the types of disfluency phenomena which all speakers exhibit to a greater or lesser extent.

Both Eklund (2004) and Roberts *et al.* (2009) note that prolongations are usually regarded as a distinctive feature of stuttered speech. However, both studies observed that relatively more speech sound prolongations occur in NF speech than had been anticipated. Eklund (2004: 280) observed that ‘for all corpora, prolongations are the third most common type, by a large margin’. He also notes that not only continuants are subject to prolongation. In the present study prolongations occur slightly more frequently than all repetitions with a range of occurrence from 1 to 9.9 prolongations per 100 syllables compared to 0.3 to 6.8 repetitions per 100 syllables.

In the present study, after filled and unfilled pauses, and prolongation, repetition is the next most common disfluency. It is hard to be certain whether an unfilled pause is deliberate or accidental therefore all pauses which meet the duration criteria were coded. In the results, unfilled pauses exhibited less speaker-specificity than other disfluency types. It is possible that greater levels of individual variation may emerge if unfilled pauses were coded with respect to location, function, etc. However, in the forensic context, applying a more fine-grained classification framework is unlikely to produce sufficient number of tokens in each category to enhance the discrimination of individuals.

While data from a larger number of speakers is needed to develop a fuller picture of population behaviour, the forensic importance of the range of disfluency types highlighted above is clear. Filled pauses offer a range of levels of occurrence for different individuals. It appears that word repetitions may be a feature for which absence/presence is an initial discriminator, with further differentiation offered by the degree of occurrence in the case of presence. For fricative prolongations, it is possible but relatively unusual for a speaker to exhibit a high rate of this feature. Observing a high rate of fricative prolongations on both questioned and suspect recordings in a forensic speaker comparison case, together with appropriate analyses of other speech variables, would therefore contribute useful information for a forensic speaker comparison report.

In addition to identifying patterns of speaker-specific behaviour in overall disfluency rates and in the occurrence of the separate types of disfluency feature, the study found considerable



individual variation in the profiles of disfluency features chosen by speakers. For example, some speakers preferred to use *er* rather than *erm*, some vice versa, and some used both types of filled pause to similar extents. Some speakers used silent pauses more often than filled pauses and vice versa. Some speakers used one or more repetition types or prolongations. The pattern of individual speakers demonstrating preferences for different types and combinations of disfluencies observed in the present study chimes with Shriberg's findings of speakers adopting different strategies in their disfluency behaviour as outlined in 1.2. In the TOFFA classification Shriberg's 'Deleters' would be speakers who tend to use interruptions.

When profiles comparing each speaker's overall selection of disfluency features in the present data were compared, marked patterns of individual behaviour were noted. This speaker-specificity was quantified using discriminant analysis such that markedly higher rates of classification were achieved for analyses combining multiple features in comparison with the separate feature analyses.

The findings presented here indicate that where possible, analysis of disfluency features should play a part in forensic speaker comparison cases. The present authors have included a broad consideration of non-fluency behaviour in a number of recent real forensic cases. In these cases, observations of patterns of disfluency behaviour provided ancillary support only for the conclusions reached, but their inclusion enhanced the overall picture. These casework results together with the findings of the present study provide a strong argument for the development of an approach to disfluency analysis grounded in population statistics which would enable quantitative evaluation of disfluency features appearing in casework recordings.

Future work must extend the analysis of disfluency features to larger groups of speakers to develop population statistics for each type of feature, towards the development of LR analyses. Further research should also explore the frequency and type of fluency features of NF speakers in different conversational contexts and speaking styles. This work will include quantification of the extent to which individual speakers are consistent in their fluency behaviour across styles, given that forensic speech recordings for comparison rarely involve the same speaking style. The way in which different accent and language groups use fluency features also requires investigation (see e.g. McDougall, Duckworth, & Hudson, 2015).

## 9 CONCLUSION

The 20 normally-fluent male speakers of SSBE examined in this study exhibited a wide range of usages of disfluencies in their speech. Speakers demonstrated extensive speaker-specific differences in their fluency profiles both in terms of the types of disfluency features they employed and their rate of occurrence. Discriminant analyses on individual rates of different types of pauses, repetitions, prolongations, interruptions and overall fluency achieved correct classifications at levels higher than chance. A discriminant analysis combining the top-performing features produced a markedly higher classification rate, demonstrating the speaker-discriminating strength of these features when considered in concert. In the context of forensic speaker comparison cases, while the results of disfluency analysis clearly cannot be relied on to any extent in isolation, such analysis is another tool in the forensic phonetician's tool box which can help contribute to the bigger picture.

## 10 ACKNOWLEDGEMENTS

The authors are grateful to the International Association for Forensic Phonetics and Acoustics for a research grant towards this study. The first author was funded by a British Academy Postdoctoral Fellowship during part of the conduct of this research. The authors are also grateful to Lucy Ellis for her involvement in a pilot study, to Anne Larsson Duckworth for her contribution to the study of inter-analyst consistency, and to two anonymous reviewers for helpful comments.

## 11 REFERENCES

- Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9, 1-26.
- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4), 351-366.
- Andrews, G., & Ingham, R. J. (1971). Stuttering: Considerations in the evaluation of treatment. *British Journal of Disorders of Communication*, 6, 129-138.
- Blankenship, J., & Kay, C. (1964). Hesitation phenomena in English speech: a study in distribution. *Word*, 20, 360-372.
- Boersma, P., & Weenink, D. (1992-2016). *Praat: A System for Doing Phonetics by Computer [computer program]*. Available at: <<http://www.praat.org/>>.
- Braun, A., & Rosin, A. (2015). In The Scientific Committee for ICPhS 2015 (Ed.), On the speaker-specificity of hesitation markers. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow 10-14 August 2015, The University of Glasgow, Paper number 731.731-734 <http://www.icphs2015.info/pdfs/Papers/ICPHS0731.pdf>.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44(2), 274-296.
- Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Language Production Volume 1: Speech and Talk*. London: Academic Press.
- Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: hesitations in speech affect language comprehension. *Cognition*, 105(3), 658-668.
- Eklund, R. (2004). *Disfluency in Swedish Human-Human and Human-machine Travel Booking Dialogues*. Linköping University Studies in Science and Technology Dissertation No. 882 [http://www.ida.liu.se/~robek28/pdf/Eklund\\_2004\\_PhD\\_Thesis\\_Corrected.pdf](http://www.ida.liu.se/~robek28/pdf/Eklund_2004_PhD_Thesis_Corrected.pdf).
- Foulkes, P., & French, P. (2012). Forensic speaker comparison: a linguistic-acoustic perspective. In P. Tiersma & L. Solan (Eds.), *Oxford Handbook of Language and Law* (pp. 557-572). Oxford: Oxford University Press.
- Fox Tree, J. E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition*, 29, 320-332.
- Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: effects of filled pauses on recall. *Journal of Memory and Language*, 2011, 161-175.
- Gilquin, G., & De Cock, S. (2011). Errors and disfluencies in spoken corpora. *International Journal of Corpus Linguistics*, 16(2), 141-172.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.
- Goldman Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech*, 4, 18-26.
- Hudson, T., de Jong, G., McDougall, K., Harrison, P., & Nolan, F. (2007). In J. Trouvain & W. J. Barry (Eds.), F0 statistics for 100 young male speakers of Standard Southern British English.

- Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken 6-10 August 2007, Universitaet des Saarlandes, 1809-1812.  
<http://www.icphs2007.de/conference/Papers/1570/1570.pdf>.
- Hughes, V. (2014). *The definition of the relevant population and the collection of data for likelihood ratio-based forensic voice comparison*. (Ph.D. Dissertation), University of York.
- 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. doi:10.1558/ijssl.v23i1.29874
- Ishihara, S., & Kinoshita, Y. (2008) How many do we need? Exploration of the population size effect on the performance of forensic speaker classification. *9th Annual Conference of the International Speech Communication Association (Interspeech)*, Brisbane, Australia, 1941–1944.
- Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, 2(4), 671–711.
- Johnson, W. (1961). *Stuttering and What You Can Do About It*. Minneapolis: University of Minnesota Press.
- Johnson, W., Darley, F., & Spriestersbach, D. (1963). *Diagnostic Methods in Speech Pathology*. New York: Harper Row.
- Kirsner, K., Dunn, J., & Hird, K. (2003) Fluency: Time for a paradigm shift. *Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5-8 September 2003, Göteborg University, Sweden*. In R. Eklund (ed.) *Gothenburg Papers in Theoretical Linguistics* 90, 13-16.
- Kjellmer, G. (2003). Hesitation. In defence of ER and ERM. *English Studies*, 84, 170-198.
- Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, 46, 117-125.
- Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics*, 4(1), 48-83.
- MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, 48(14), 3982-3992.
- McDougall, K. (2004). Speaker-specific formant dynamics: an experiment on Australian English /aɪ/. *International Journal of Speech, Language and the Law*, 11(1), 103-130.
- McDougall, K. (2006). Dynamic features of speech and the characterisation of speakers: towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law*, 13(1), 89-126.
- McDougall, K., Duckworth, M., & Hudson, T. (2015). In The Scientific Committee for ICPhS 2015 (Ed.), Individual and group variation in disfluency features: a cross-accent investigation. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow 10-14 August 2014, University of Glasgow, Paper number 0308.0301-0305.  
<http://www.icphs.info/pdfs/Papers/ICPHS0308.pdf>
- Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law*, 16(1), 31-57.
- Roberts, P. M., Meltzer, A., & Wilding, J. (2009). Disfluencies in non-stuttering adults across sample lengths and topics. *Journal of Communication Disorders*, 42, 414-427.
- Schiel, F., & Heinrich, C. (2015). Disfluencies in the speech of intoxicated speakers. *International Journal of Speech Language and the Law*, 22(1), 19-33. doi:10.1558/ijssl.v22i1.24767
- Shriberg, E. (2001). To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31, 153-169.
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using Multivariate Statistics* (6th ed.). Harlow: Pearson.
- Van Riper, C. (1973). *The Treatment of Stuttering*. Englewood Cliffs, NJ: Prentice-Hall.
- Wingate, M. E. (1964). A standard definition of stuttering. *Journal of Speech and Hearing Disorders*, 29, 484-489.