INDIVIDUAL AND GROUP VARIATION IN DISFLUENCY FEATURES: A CROSS-ACCENT INVESTIGATION

Kirsty McDougall,[1] Martin Duckworth[2] and Toby Hudson[1]

[1]Phonetics Laboratory, Department of Theoretical and Applied Linguistics, University of Cambridge;
[2]Duckworth Consultancy Ltd
kem37@cam.ac.uk, msd@duckworth-consultancy.co.uk, toh22@cam.ac.uk

ABSTRACT

A study of individual differences in the fluency disruptions of speakers of two different accents, Standard Southern British English (SSBE) and York English is presented. Distributions of rates of occurrence per 100 syllables are examined for filled and silent pauses, repetitions, prolongations and (self-)interruptions, and subcategories of these. Patterns of occurrence of disfluency features show considerable between-speaker variation in both SSBE and York English. Similar ranges of speakers' overall disfluency rates are exhibited by both accents, but cross-accent differences are present in the patterning of some disfluency feature categories. The results suggest that a detailed record of disfluency features is a useful additional tool in forensic speaker comparison.

Keywords: disfluency, individual differences, accent differences, forensic speaker comparison

1. INTRODUCTION

Features of speech related to fluency such as silent and filled pauses, sound prolongations, repetitions and self-interruptions exhibit variation between speakers. However, the range of individual variation in non-fluency phenomena in speakers identified as having normally fluent speech has received little attention in phonetic research. Non-fluency phenomena have clear potential for individual variation since they are likely to play a part in the planning of speech (e.g. [3]), may be influenced by psycho- or socio-linguistic demands, and are difficult – maybe even impossible – to control consciously.

Analysis of individuals' use of disfluencies has great potential for application in forensic speaker comparison cases, in which voice recordings of an unknown speaker committing a crime and a suspect are compared. The bulk of the literature on speaker-distinguishing properties of speech for forensic applications has focussed on phonetic features which bear a direct relationship with a speaker's anatomy, e.g. mean f0 and formant frequencies (e.g. [3]). Investigating the speaker-distinguishing potential of disfluency features focusses on a very different aspect of a speakers' performance: speech features which are behavioural rather than anatomical. Further, disfluency features are largely realised through the temporal domain and therefore generally well-preserved in the poor recording conditions of forensic cases where telephone transmission and background noise are typical. This is in contrast to the 'anatomical' features mentioned above which are conveyed through spectral information for which adverse recording conditions are more problematic.

Developing an understanding of individual variation in normal non-fluency behaviour is also of importance in speech and language therapy where knowledge of the extent of disfluency behaviours among speakers with no speech or hearing problems provides a reference point for therapists working with stammering speakers. Although the study by Roberts *et al.* [7] on non-stuttering adults found relatively few disfluencies in their sample, the analysts counted some phenomena only if they sounded like stuttering. Eklund [1] on the other hand used a different metric and found that non-stuttering speakers produced a larger number of disfluencies. The taxonomy and methods of counting disfluencies require careful definition.

[11] presents a study of individual behaviour in the fluency disruptions of 20 male SSBE speakers from the *DyViS* database [6] undertaking a simulated police interview and a telephone call with a 'friend'. The rate of occurrence per 100 syllables of a range of disfluency features was calculated for each speaker in the two speaking styles. Results showed that individuals varied considerably in their rates of fluency disruptions and that individual differences were present in their 'disfluency profiles', i.e. the types of disfluencies each speaker produced. Disfluency features also showed a degree of within-speaker consistency across the two speaking styles.

The present study investigates whether patterns of disfluency differ across different accent/cultural backgrounds, by comparing the disfluency behaviour of speakers of SSBE and of York English. The study considers whether the range of individual variation observed in SSBE mirrors that observed in York English, whether disfluency features could be considered accent-independent to any extent, and the implications of these findings for forensic phonetics.

2. METHOD

2.1. Speakers and recordings

20 male speakers of SSBE, aged 18-25, were selected from the *DyViS* database [6]. Recordings analysed were mock police interviews where the participant was in the role of 'suspect' and responded to interview questions based on map-task style materials (*DyViS* Task 1). 20 male speakers of York English, also aged 18-25 years, were recorded undertaking the same tasks as in *DyViS* to create the 'YorViS' database. Task 1 recordings from *YorViS* were also analysed. The mean duration of the interviews was 14 mins in *DyViS* and 15 mins in *YorViS*.

2.2. Taxonomy of disfluency feature categories

The study adopted a general definition of a 'fluency disruption' as follows: *any phenomenon originated by the speaker which changes the flow of the speaker's utterance.* A system for classifying the speech disfluencies in each recording was devised combining features of Shriberg's taxonomy for non-pathological speech [8] and those of Wingate [10] and Van Riper [9] for pathological speech. A brief outline of the disfluency categories used is given in Table 1 (see [11] for a detailed outline).

2.3. Annotating transcriptions and tabulating results

The speech data were transcribed orthographically in *Praat* text grids, with the disfluency features annotated. The transcriptions were transferred to a spreadsheet, where the number of phonetic syllables per utterance was also recorded. The syllables counted included all repetitions, even part-word repetitions, but excluded all non-word phenomena such as filled pauses. The number of occurrences of each disfluency feature per 100 syllables was calculated for each speaker. This was carried out for 1000 syllables per speaker, where available (for 5 SSBE speakers and 5 York speakers 800-990 syllables were analysed).

3. RESULTS

3.1. All disfluencies

Histograms showing the distribution of individuals' overall rates of production of disfluency features are shown in Figure 1, with a separate panel for SSBE and York English.

In both accents, a wide range of individual variation is present, with SSBE speakers ranging between 13.0 and 36.8 disfluencies per 100 syllables and York speakers ranging between 9.4 and 32.9.

Table 1: Categories of disfluency features.

	Subcategories and examples
Filled Pauses	- *er* - *erm* - others, e.g. *ah*
Silent Pauses	- 'grammatical' - 'other'
Repetitions	- part-word *on the road I park my car th-there's* - whole word *but she- she's also* - phrase *on your-on your left there's a reservoir* - multiple (i.e. more than 2 iterations) *a hairdresser at the- at the- at the- at the-*
Prolongations	(duration ≥ 200 msecs) - vocalic, e.g. vowel, nasal, lateral - fricative - plosive closure duration or affricate closure or release duration
Interruptions	(speaker interrupts self and discontinues the utterance, or continues with a modification) - phrase *pighty road which- and then then you ...* - word *I th- I probably recognise like the bar lady*

Figure 1: Individuals' overall rates of disfluencies in SSBE and York English.
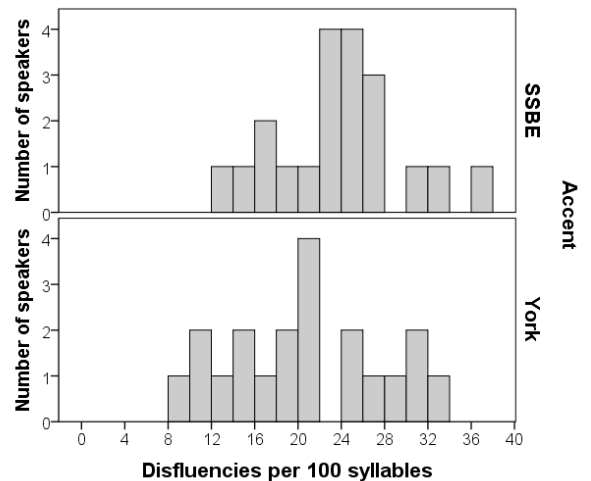
The pattern of SSBE speakers exhibiting a greater overall rate of disfluencies is not confirmed statistically: a Mann-Whitney test shows no significant difference between SSBE (*Mdn* = 23.64) and York (*Mdn* = 20.65) ($U = 155.0$, $p = 0.229$).

3.2. Filled and silent pauses

Histograms showing each speaker's rate of production of filled pauses (*er, erm* and 'other'

pauses combined) for each accent are shown in Figure 2. SSBE speakers (*Mdn* = 8.46) produce more filled pauses than York speakers (*Mdn* = 5.90) as is clear from the histograms and confirmed by a Mann-Whitney test (*U* = 87.50, *p* = 0.002). Examining the results for the subcategories of filled pauses shows that *er* appears to be the main contributor to the difference between the accents, with SSBE speakers also producing significantly more *er* pauses (*Mdn* = 5.01) than York speakers (*Mdn* = 3.15) (*U* = 115.00, *p* = 0.021).

By contrast, rates of silent pausing do not exhibit differences between the two accents (*U* = 136.50, *p* = 0.087)

3.3. Repetitions

The occurrences of repetition per 100 syllables (part-word, whole word, phrase and multiple, combined) do not differ significantly between the two accents (*U* = 140.00, *p* = 0.106). However, inspecting the results for the separate subcategories reveals a significant difference in the rates of word repetition between the two accents, with SSBE (*Mdn* = 1.20) producing higher rates than York (*Mdn* = 0.41) (*U* = 112.50, *p* = 0.017), as shown in Figure 3.

3.4. Prolongations

Individuals' rates of prolongations (vocalic, fricative and plosive/affricate, combined) in the two accents are shown in the histograms in Figure 4. For this feature, York speakers (*Mdn* = 5.00) are more prolific per 100 syllables than SSBE (*Mdn* = 2.70) as is confirmed by a Mann-Whitney test (*U* = 112.00, *p* = 0.016). An examination of the separate subcategories shows that both vocalic and fricative prolongations are contributing to this difference, with York speakers (*Mdn* = 2.35) producing significantly more vocalic prolongations per 100 syllables than SSBE speakers (*Mdn* = 1.35) (*U* = 90.50, *p* = 0.002) and significantly more fricative prolongations per 100 syllables (York *Mdn* = 1.26, SSBE *Mdn* = 0.60) (*U* = 114.0, *p* = 0.019).

3.5. Interruptions

Histograms showing speakers' rates of interruptions (word and phrase combined) in the two accents are given in Figure 5. This is the second category of disfluencies in which York speakers produce higher rates than SSBE, a Mann-Whitney test showing a significant difference between York (*Mdn* = 1.66) and SSBE (*Mdn* = 1.11), albeit with a relatively large *p*-value (*U* = 127.00, *p* = 0.048). Perusal of the

Figure 2: Individuals' rates of occurrence of filled pauses in SSBE and York English.
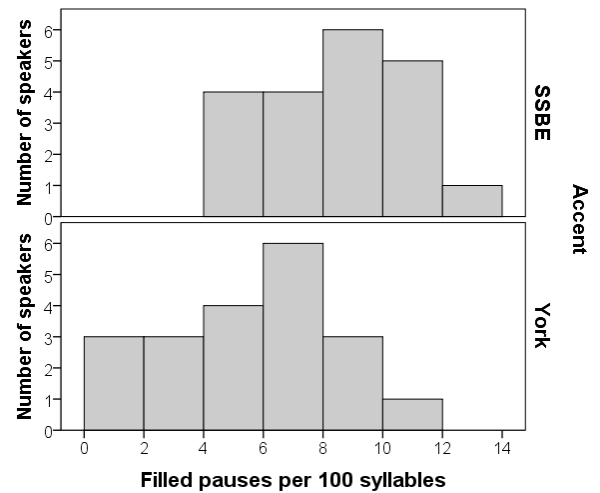

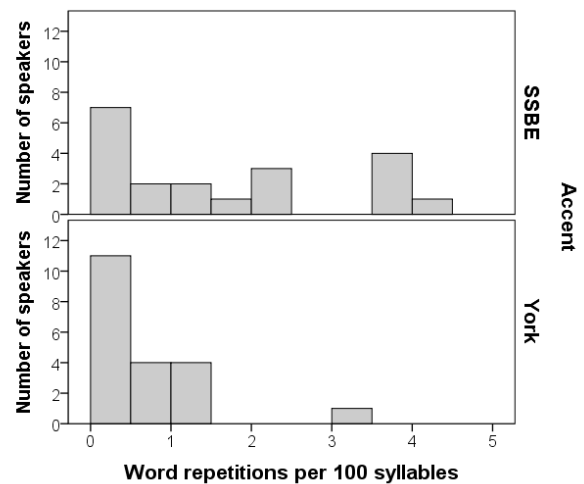**Figure 3**: Individuals' rates of occurrence of word repetitions in SSBE and York English.

Figure 4: Individuals' rates of occurrence of all prolongations (vocalic, fricative, plosive/affricate) in SSBE and York English.
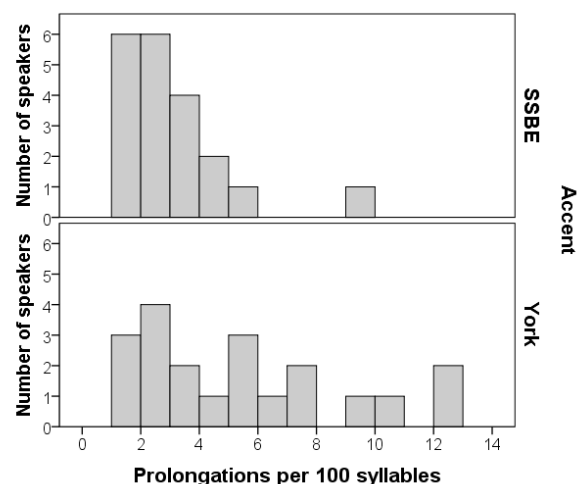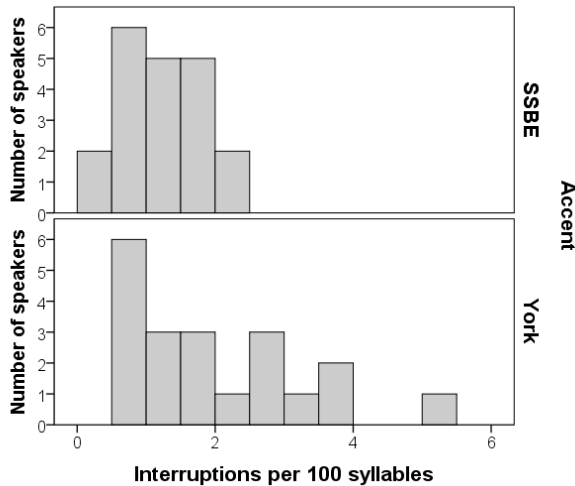
Figure 5: Individuals' rates of occurrence of all interruptions (phrase, word) in SSBE and York English.

subcategory results shows that word interruptions bear the same direction of difference (SSBE $Mdn = 0.021$, York $Mdn = 0.59$; $U = 87.50$, $p = 0.002$) while phrase interruptions show no significant differences between the accents ($U = 151.50$, $p = 0.194$).

4. DISCUSSION

Comparing the overall distributions, it is clear that disfluency rates show considerable variation among individuals, and that this large variation is present in both accents. Both distributions appear to be approximately normally distributed such that many speakers' rates are located in the histogram bars centred around each mean. However, for speakers with overall disfluency rates towards the tails of each distribution, this feature may be relatively distinctive. In other words, the data presented here form the beginnings of a background reference population for disfluency features, analogous to e.g. [5] and [4] for long-term f0 in German and SSBE respectively, from which forensic phonetic inferences about how typical an individual's performance is for a particular speech feature may be drawn.

Patterns of overall disfluency rate were similar across the two accents, lending some support to the possibility that disfluencies may have a similar range of occurrence regardless of accent. However, when the detail of the separate disfluency feature categories is examined, some accent differences emerge. SSBE speakers produce filled pauses more frequently than York speakers, and in particular more *er* pauses. SSBE speakers also produce more word repetitions. By contrast, SSBE has fewer prolongations and fewer interruptions than York

English. Given the extent of individual differences also observed for these features, a larger data set is required to draw stronger conclusions, yet it appears that the profile of disfluency features used by speakers can differ across accents.

The speakers are likely to differ in educational and cultural background, in addition to accent. The SSBE speakers were students of the University of Cambridge, while the York speakers were recruited from public locations, e.g. outside a job centre. Some of the York speakers were less forthcoming in their interviews than the SSBE speakers, and several also demonstrated difficulties in reading some of the materials. Thus the two speaker groups studied differed not only by accent, but also by socio-cultural background, both of which may affect a range of speech phenomena including f0, voice quality, speaking rate and fluency. There may however be something speaker-specific about disfluency: it serves a cognitive function (speech planning, speech decoding) but there may also be something distinctive (a learned behaviour) which may or may not be related to a regional marker like accent.

5. CONCLUSION

Disfluency features provide an interesting source of individual variation, useful as a complement to other speech features analysed in forensic speaker comparison. The data for SSBE and York English examined here showed similar levels of overall disfluency across speaker groups, but group-specific differences in certain disfluency subcategories were present. Further work is needed to examine the individual disfluency profiles of the York speakers to determine whether the extent of speaker-discriminating information exhibited by the individual profiles of SSBE speakers [11] is also present. Future work with larger speaker groups and for additional accent/cultural groups is needed to determine the extent to which the patterns of group and individual behaviour observed here apply more widely.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] 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.divaportal.org/smash/get/diva2:20923/FULLTEXT01.pdf&sa=U&ei=4zJaU5izEufL2QWI3IGoBA&ved=0CCAQFjAB&usg=AFQjCNH-uRPS0AkzYNxOjMQoaXgSZ3SIOA, accessed 21/01/2015]

[2] Foulkes, P., French, P. 2012. Forensic speaker comparison: a linguistic-acoustic perspective. In: Tiersma, P., Solan, L. (eds), *Oxford Handbook of Language and Law*. Oxford: Oxford University Press, 557-572.

[3] Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in Spontaneous Speech* London: Academic Press.

[4] Hudson, T., de Jong, G., McDougall, K., Harrison, P., Nolan, F. 2007. F0 statistics for 100 young male speakers of Standard Southern British English. *Proceedings of the 16th International Congress of Phonetic Sciences.* Saarbrücken, Universitaet des Saarlandes.

[5] Künzel, H. J. (1989): How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, 46, 117-125.

[6] 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.

[7] 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.

[8] Shriberg, E. (2001) To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31, 153-169.

[9] Van Riper, C. 1982. *The Nature of Stuttering.* (2nd ed.) Englewood Cliffs, N.J.: Prentice-Hall.

[10] Wingate, M. E. 1964. A standard definition of stuttering. *Journal of Speech and Hearing Disorders* 29, 484-489.

[11] Duckworth, M., McDougall, K. In prep. Individual differences in normal non-fluency in adult males: a study of Standard Southern British English.