APPLICATION OF THE ‘TOFFA’ FRAMEWORK TO THE ANALYSIS OF DISFLUENCIES IN FORENSIC PHONETIC CASEWORK

Kirsty McDougall[1], Richard Rhodes[2,3], Martin Duckworth[4], Peter French[2,3] and Christin Kirchhübel[5]

[1]University of Cambridge, [2]J P French Associates, [3]University of York, [4]Independent Researcher,
[5]Soundscape Voice Evidence, formerly at J P French Associates
kem37@cam.ac.uk, richard.rhodes|peter.french@jpfrench.com, martinsduckworth@gmail.com, ck@soundscapevoice.com

ABSTRACT

Although disfluency features such as filled and silent pauses, repetitions, prolongations and self-interruptions can be expected to exhibit a range of individual variation, until recently there was little research quantifying this variation for normally-fluent speakers. Previous analysis of disfluencies in forensic speaker comparison (FSC) casework had been limited to impressionistic description rather than analysis within a quantitative framework. In 2017, McDougall and Duckworth published TOFFA, a ‘Taxonomy of Fluency features for Forensic Analysis’ [5], which provides a formal system for quantifying individual variation in normally-fluent speakers in the forensic context. The present paper discusses points to consider in implementing the framework in casework. Example cases from the experience of the consultancy J P French Associates (JPFA) are presented to illustrate situations where analysis of disfluencies was of key importance. The work provides evidence that in cases where it can be used, disfluency profiling using TOFFA is a useful tool for FSC, complementary to other types of analysis.

Keywords: fluency behaviour, disfluency features TOFFA, individual differences, speaker-specificity.

1. INTRODUCTION

Normally fluent speakers are not perfectly fluent. Their speech exhibits a range of perturbations to its flow such as filled and silent pauses, repetitions, prolongations and self-interruptions. Such phenomena are of interest in forensic phonetics due to their potential for individual variation. Usage of filled and silent pauses may play a part in the planning of speech and is therefore likely to differ among speakers. Variation between speakers can likewise be expected for fluency disruptions such as repetitions and prolongations which are also related to the planning and execution of speech processes, and therefore difficult to control consciously or exploit for disguise.

In a FSC case, the phonetician is asked to compare a speech recording related to a crime with that of a known suspect, with a view to assessing the likelihood that the speech of the same speaker is present in both recordings. Depending on availability of material, this analysis will typically include the comparison of auditory observations and acoustic measurements in respect of a range of phonetic features; many of these are related to the anatomy of the speaker, e.g. fundamental frequency and formant frequencies. Analysing disfluency features allows the phonetician access to a source of behaviour-related information about a speaker which is complementary to the information provided by anatomy-related features. Further, while many anatomy-related features are realised in the spectral domain, disfluency features are primarily realised in the temporal domain so are less susceptible to the detrimental effects of telephone transmission at play in a considerable proportion of forensic recordings.

Until recently, analysis of disfluency behaviour in forensic casework had been a matter of ad hoc description rather than analysis within a framework of formal categories. With two notable exceptions [2,3], quantitative investigation of disfluency for potential forensic application has been limited.

In [6], an approach for quantifying the disfluency behaviour of individuals is described using TOFFA: ‘Taxonomy of Fluency features for Forensic Analysis’. TOFFA draws on ideas developed in research analysing features of the speech of people who stutter, e.g. [4,10]. Using TOFFA, McDougall and Duckworth analysed the disfluency profiles of 20 male speakers of Standard Southern British English (SSBE), aged 18-25 years, in interview-style speech and telephone conversations from the *DyViS* database [8]. The 20 speakers displayed considerable individual variation in their disfluency behaviour, both in terms of the types of disfluency features they used and in their rates of occurrence [5]. When comparing across the two speech styles, individuals’ disfluency profiles showed a degree of consistency for most disfluency features [6]. TOFFA has also been used to demonstrate patterns of individual variation in the disfluency profiles of 20 male speakers of York English [7]. While the speaker-specificity demonstrated by the TOFFA disfluency profiles of the 20 SSBE and 20 York English speakers is very encouraging, larger-scale empirical data is not yet available for different types of speakers, nor for intra-speaker variability across a wider range of communicative situations; there is a great deal of scope for future research in this area.

The development of McDougall and Duckworth's taxonomy has proceeded symbiotically through theoretical work in the laboratory, their own casework investigations and regular discussion with other forensic phoneticians, including the co-authors of the present paper. This paper thus gives a brief outline of the TOFFA approach followed by a discussion of its implementation in casework. Pertinent findings and issues from three criminal cases are presented. Implications of this work for future research and ongoing practical concerns in enhancing the forensic application of TOFFA are discussed.

2. THE 'TOFFA' FRAMEWORK AND METHOD

The TOFFA approach adopts a general definition of a 'fluency disruption' as: *any phenomenon originated by the speaker which changes the flow of the speaker's utterance.* A brief outline of the disfluency categories used is given in Table 1 (see [5] for a detailed outline).

To produce a TOFFA profile, speech data are transcribed orthographically in a text grid and the disfluency features annotated, using software such as *Praat* [1]. The transcriptions are transferred to a spreadsheet, along with a record of the number of phonetic syllables per utterance, or a note of the duration of the utterance. The syllables counted include all repetitions, even part-word repetitions, but exclude non-word phenomena such as filled pauses. The number of occurrences of each disfluency feature per 100 syllables or per unit time-stretch is calculated for each speaker.

3. USING TOFFA IN CASEWORK

Previous examination of disfluencies for FSC cases was typically based on impressionistic description, so lacking in quantification. It was therefore difficult to assess the evidential value of disfluencies, e.g. using an approach such as likelihood ratios. Judgements that forensic phoneticians made about the typicality of certain disfluency features were purely subjective, by necessity given the lack of population data or a formal quantitative framework for analysis. Reports tended only to mention disfluency for speakers where disfluency patterns were very marked, e.g. extreme stuttering or very frequent use of a particular feature. In such cases disfluency behaviour was only described at an impressionistic level. An example of the kind of wording that was used is: *"The speech of the suspect and the offender featured a very similar pattern of disfluency. There was evidence of stammering in both recordings, which manifested in frequent repetitions of words and syllables, and block-type fluency interruptions."*.

The TOFFA framework offers a more objective approach with a clear methodology, enabling precise quantification and replication of findings. Further, TOFFA enables the analyst to capture features that are not necessarily perceptually salient.

The forensic consultant co-authors have in recent years adapted the TOFFA approach at JPFA to meet the practical needs of the cases in which they have included disfluency analysis. While McDougall and Duckworth [5] propose counting the number of occurrences of each particular disfluency type against a syllable base (number of occurrences per 100 syllables), the JPFA co-authors pointed out that using metrics anchored within a time-base would increase efficiency since the data counts could be collected in parallel with the analysis of other forensically-relevant features. This prompted a further study by McDougall and Duckworth [5] comparing the levels of speaker-discriminating information yielded by the same data set (the 20 SSBE speakers from *DyViS* in the police interview task) using a time-base and a syllable-base. The time

Table 1: Categories of disfluency features.

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

-based metrics were the number of occurrences of the feature of interest per 20s stretch of speech (edited compilation of the target speaker's speech). High correlations ($r \geq 0.8$) were achieved between syllable- and time-based measures. Discriminant analyses showed small differences in classification accuracy for each approach, but neither was consistently better. Since the two approaches provided similar levels of speaker-specific information on the whole, the first and third author have proceeded with a time-based approach in subsequent research and JPFA have used time-based analyses in casework.

Two simplifications to TOFFA categories have been adopted for the JPFA implementation to casework. Firstly, while TOFFA captures both pauses at grammatical boundaries **[pg]** and pauses in other locations within an utterance **[po]**, the JPFA implementation counts **[po]** only; **[pg]** is not among the features analysed. Secondly, while TOFFA differentiates between two types of consonantal prolongations **[prof]** and **[prop]**, the JPFA implementation merges these into one consonant category, **[proc]**.

Measuring disfluency behaviour is not appropriate in every FSC case. First of all, the samples need to be long enough to establish meaningful rates of disfluency. Whether a sample is sufficiently long depends on the quantity of speech and on the type of content, i.e. natural flowing speech versus giving addresses, telephone numbers, etc. At present, JPFA would be unlikely to undertake a quantitative TOFFA analysis unless around 45-60s net speech is available in both the known and questioned samples. A goal for further research is to improve our understanding of how much speech is required to establish stable disfluency rates, and the impact of various relevant factors. Secondly, given the time implications, currently a TOFFA analysis is more likely to go ahead for a forensic case where it appears impressionistically that disfluency profiling would be helpful. Further considerations about the comparability of the speaking styles and situations of the recordings to be compared come into play. While the degree of (mis)match between the technical recording characteristics of the samples can vary, it is desirable for there to be a degree of similarity in the power relations, cognitive load, interlocutor, topic and guise across the two recordings. As mentioned in the Introduction, a lack of empirical data on the effects of these factors or the extent of intra-speaker variation across different real world communicative situations means that it is important that the limitations of any conclusions drawn using TOFFA are carefully stated.

The JPFA authors note that specific training is needed to become skilled in using TOFFA. They emphasise the importance of collaborative work – checking analyses jointly and monitoring cross-calibration of experts, since there is a degree of subjective judgment in assigning features to categories.

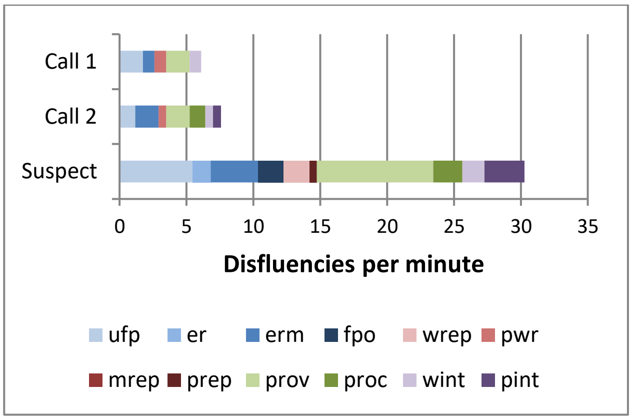
4. CASEWORK EXAMPLES

Some examples of the application of disfluency analysis using TOFFA to cases are presented below.

4.1. Case 1

The graph in Figure 1 shows the disfluency profiles yielded by two phone calls made by (an) unknown speaker(s) and the speech sample of a suspect in a fraud case. For these recordings, the TOFFA analysis added further strength to other findings supporting the hypothesis that different speakers were involved, as the suspect sample was considerably more disfluent (30 disfl/min) than either of the questioned call samples (Call 1: 6 disfl/min; Call 2: 8 disfl/min). The TOFFA results were consistent with differences between the recordings in voice quality, accent, vowel formants, rhythm and grammatical features.

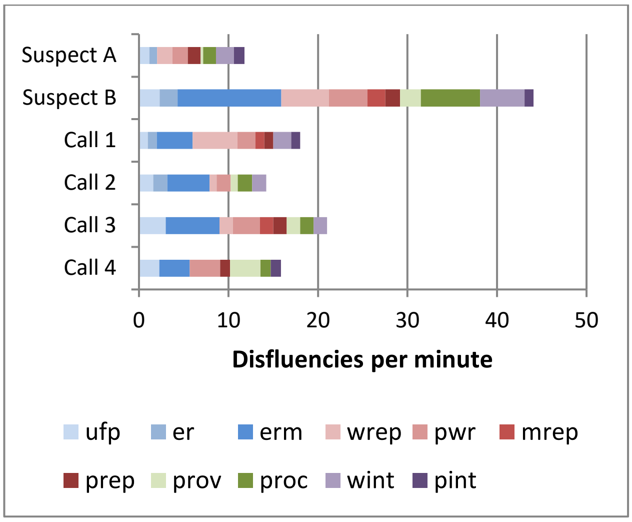
Figure 1: Case 1 TOFFA disfluency profiles for Call 1 and Call 2 (unknown speaker) and suspect sample.



4.2. Case 2

In this case, the task was to compare several incriminating phone calls with two suspects who were brothers. The two suspect voices were extremely similar; they yielded similar measures for pitch, voice quality, overall accent features, and F1, F2 and F3 frequency distributions. Although there were some segmental variations and differences in rhythmic behaviour between the brothers, these were slight. TOFFA profiling (Figure 2), however, contributed considerable speaker-distinguishing information, such that Suspect B's speech was much more disfluent (44.1 disfl/min) than Suspect A's

Figure 2: Case 2 TOFFA disfluency profiles for Call 1 and Call 2 (unknown speaker) and suspect sample.



(11.8 disfl/min). The overall disfluency rates presented by the speaker in the four phone calls were 14.2, 15.8, 18.0 and 21.0 disfl/min, i.e. the offender speech was considerably less disfluent than Suspect B, and was more similar to Suspect A.

4.3. Case 3

This case centred on a fraudulent phone call in which a young adult woman disguised her voice to impersonate an elderly man. The disguise involved lowering of the larynx, such that many of the analyses usually undertaken for FSC - such as the evaluation of formant frequencies, pitch and voice quality features - were potentially unreliable. However, although empirical data is not yet available, one might hypothesise that for certain types of disguise, disfluency features may be unaffected. In the case presented here, the suspect and offender samples yielded similar rates and types of disfluencies in their TOFFA profiles, as shown in Figure 3 (although more prolongations were present in the interview). The extent of consistency of disfluency profiles across disguised and undisguised speech is a topic requiring further research.

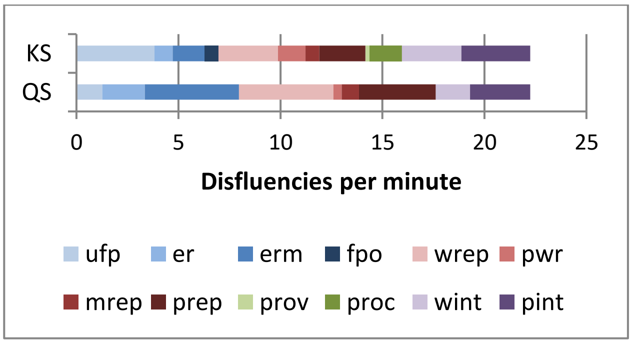
5. DISCUSSION

The quantitative footing now given to the analysis of disfluency behaviour by TOFFA is an important step forward, but there is still a long way to go. The three casework examples presented here show quantification of interesting disfluency usage, yet what is missing is population data for TOFFA profiles to enable the assessment of the typicality of a particular disfluency profile against the backdrop of an appropriate population distribution. As is the case for the majority of features examined by forensic phoneticians using an auditory/acoustic phonetic approach, at present, analysts are only able to make typicality judgements for disfluency features based on casework experience, i.e. the gradual development of a 'mental database' by carrying out TOFFA analysis on previous cases. Judgements therefore remain subjective, albeit on a more quantitatively replicable set of results.

The TOFFA approach, its early lab-based research results [5,6,7] and the illustration of its implementation to casework data here, provide a firm foundation for the next stage of improving the forensic analysis of disfluency behaviour. Further research is needed to determine the distribution of disfluency features across larger normally-fluent populations, across a range of different speaking situation, across non-contemporaneous recordings, and across different accents and varieties. When such data become available, analyses testing the speaker-specificity of TOFFA disfluency profiling using likelihood ratio analyses should be undertaken [9]. In the meantime, some progress could be made within the casework context, by checking that analysts are measuring in the same way and then using case data to compile a database of reference material towards improved typicality assessments. In fact, JPFA are in the process of compiling disfluency profile data - along with data for a range of other features - from case files in order to generate this type of resource.

While population-based calculations of the speaker-specificity of disfluency features remain a longer-term goal, TOFFA does offer quantification of a behavioural aspect of speech which previously was only analysed in impressionistic terms for forensic cases. The fact that disfluency features are generally well-preserved in forensic recordings and are a complementary source of information about a speaker makes them a very attractive domain for further research and application in forensic work.

Figure 3: Case 3 TOFFA disfluency profiles for the known sample (KS) and questioned (disguised) sample (QS).



6. REFERENCES

[1] Boersma, P. and Weenink, D. 1992-2018. *Praat: A System for Doing Phonetics by Computer* [computer program]. http://www.praat.org/

[2] Braun, A., Rosin, A. 2015. On the speaker-specificity of hesitation markers. *Proc. 18th ICPhS* Glasgow. Paper number 731.

[3] 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.

[4] Johnson, W., Darley, F.L., Spriestersbach, D.C. 1963. The problem of stuttering. *Diagnostic Methods in Speech Pathology*. New York: Harper and Row, Chapter 9.

[5] McDougall, K., Duckworth, M. 2017. Profiling fluency: an analysis of individual variation in disfluencies in adult males. *Speech Communication* 95, 16-27.

[6] McDougall, K., Duckworth, M. 2018. Individual patterns of disfluency across speaking styles: a forensic phonetic investigation of Standard Southern British English. *International Journal of Speech Language and the Law* 25(2), 205-230.

[7] McDougall, K., Duckworth, M., Hudson, T. 2015. Individual and group variation in disfluency features: a cross-accent investigation. *Proc. 18th ICPhS* Glasgow, Paper number 0308.

[8] 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.

[9] Rose, P. 2002. *Forensic Speaker Identification*. London: Taylor and Francis.

[10] Ward, D. 2006. The assessment and measurement of stuttering. *Stuttering and Cluttering: Frameworks for Understanding and Treatment*. Hove, East Sussex: The Psychology Press, Chapter 9.