A TECHNIQUE FOR THE PHONETIC TRANSCRIPTION OF STUTTERING

JONATHAN HARRINGTON

The Centre for Speech Technology Research and Department of Linguistics, University of Edinburgh, Scotland.

ABSTRACT

This paper describes a phonetic transcription system that has been developed from an auditory, spectrographic and electrolaryngographic analysis of the production of stuttered speech by 36 adult stutterers and an electropalatographic study of stuttered speech produced by 2 adult subjects. The system enables the transcription of both 'repetitions' and 'prolongations' and provides some guidelines for their distinction. In the final section, a description is given of some of the characteristics to which the production of stuttered speech seems to conform.

1. INTRODUCTION

One of the main advantages of the development of a phonetic transcription system for stuttering is that it enables the classification of stuttered speech which may in the past have been obscured by the usage of ill-defined terminology As Wingate [1] has noted, some of the more common terminology includes: repetitions, prolongations, interjections, part-word repetitions, word repetitions, phrase repetitions, blocks, blocking, blockades, silent blocks, hard contacts, forceful attacks, spasms, broken words, revisions and incomplete phrases. Cutting across this, there is also 'tonic' and 'clonic' stammering and stuttering and variations thereof, including 'primary clonus', 'tonoclonus', 'clonotonus' and 'initial tonus' [2] [3]. In a phonetic transcription, auditory impressions are componentially analysed in terms of a finite set of articulatory parameters with recognisable acoustic correlates: therefore, confusion which may result from the welter of labels referred to above is to a large extent eliminated, since their inclusion is not necessary. At the same time, since a phonetic transcription has articulatory referents, it can provide a convenient bridge between an auditory impression of stuttered speech and an empirical analysis using physiological or acoustic techniques.

The transcription system reported in this paper is based on the production of around 800 stuttering disfluencies produced by 36 stutterers and transcribed by a trained phonetician. A full list of the transcribed material is given in [4].

2. METHOD

36 adult, male and female stutterers all attending speech therapy clinics around Edinburgh and Cambridge were recorded in a sound treated recording studio. Two gold-plated, surface electrodes from a Fourcin electrolaryngograph were secured with a band around the subject's neck at the level of the thyroid cartilage. The electrolaryngographic signal was stored on channel 2 of the Revox A77, channel 1 being used for the audio signal. The subject read the first two paragraphs of the *Rainbow Passage* and avoided, as far as possible, the use of any 'techniques' to improve fluency which may have been learned at speech therapy clinics. Following the *Rainbow Passage*, the subject relaxed for at least five minutes while the next recording was prepared in which the experimental design was the same as above. In addition, connecting wires were fed from a Tektronix TM 504 frequency generator to a pair of headphones worn by the subject in the recording studio. The subject was asked to read a list of 200 monosyllabic words one at a time following the offset of a 1 kHz tone from the frequency generator. The stimulus was activated by the experimenter following the production of each word by the subject and was designed to prevent coarticulation across word boundaries.

For the electropalatographic recording, two subjects, one male, one female, were selected from the population of 36 stutterers. In choosing the subjects, it was necessary to ensure that the majority of their disfluencies were realised as some form of lingual-palatal contact. An upper plaster cast impression was made for each subject and from this an acrylic palate containing 62 silver electrodes extending as far back as the junction of the hard and soft palate. These two subjects read the same list of 200 monosyllables described above; in addition, a Reading University electropalatograph [5] connected to a Commodore 3032 computer was used at a sampling rate of 100 Hz to store the palatograms as a function of time.

3. PHONETIC TRANSCRIPTION SYSTEM

3. 1 Syntagmatic Analysis

A subsequent analysis of the transcription suggested that disfluencies can be classified into four broad types:

(1) $\overbrace{[\text{e}]}^{\{\ \}}$ [min] (2) [m] [m] [min]

(3) [m m min] (4) $[\overline{\text{m}\qquad}\text{min}]$

The first type of disfluency, which does not bear any phonetic relationship to the target syllable (*mean*), is sometimes referred to as an *interjection*: the production of pause filling sounds such as 'er', 'um' are included under this category. When such interjections are produced, they have been transcribed phonetically (as far as this is possible), and a brace notation has been placed above the transcription. In (2), the disfluency is realised as two [m] segments.

Two criteria have been adhered to in this type of transcription. First, the duration of the [m] segments of the disfluency is approximately the same as the duration of [m] in the target syllable. Second, after the production of each [m] of the disfluency, there may be a pause of several seconds, in which the subject might exhale and inhale. If this kind of transcription has been used, the clear segmentation of the separate [m] segments using acoustic techniques should always be possible. In (3), the [m] segments are also of approximately the same duration as [m] of the target syllable but they are transcribed inside the brackets to indicate that the duration between them is negligible. Spectrographic and laryngographic analyses of type (3) disfluencies show a succession of similar 'events' which often cannot easily be segmented. In the laryngogram of the disfluency in Figure 1, realised as [n n n n n n n], it is possible to detect by eye the repetition of a similar pattern 8 times, but the boundaries between successive repetitions cannot be easily determined. Another characteristic

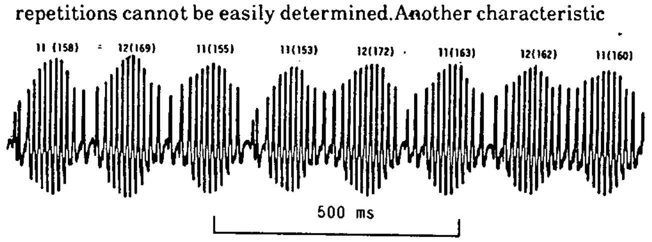

FIGURE 1: Laryngogram of a series of [n] repetitions (target syllable *no*). The estimated number of periods for each repetition is shown above the laryngogram; the number in brackets, adjacent, shows the corresponding duration in milliseconds.

feature of type (3) disfluencies is that the duration of the repeated segments tends to be approximately equal. In Figure 1 for example, each 'segment' appears to consist of either 11 or 12 cycles. In the spectrogram in Figures 2 of a disfluency realised as [pʰ pʰ pʰ pʰ pʰ] (target syllable *piece*) the duration of successive [pʰ] segments varies betweeen 172 ms and 185 ms.

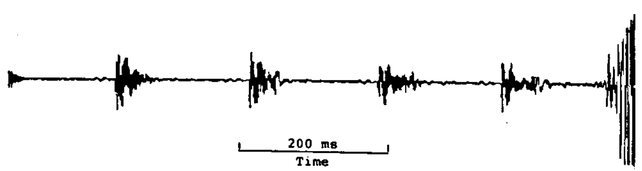

FIGURE 2: Audio wave (digitised at 16 kHz, low-pass filtered to 8 kHz in AUDLAB[1] on a MassComp MC-500) of four [pʰ] segments (target syllable *piece*). The durations of the four [pʰ] segments (closure and aspiration) are 172 ms, 181 ms, 185 ms and 172 ms respectively.

Unlike type (2) disfluencies, the vocal organs do not return to a neutral position between successive segments.

A transcription of type (4) has been used when the disfluency is realised as a prolonged section of the prevocalic consonant(s) of the target syllable and is continuous with the vowel: the duration for which the segment is produced is indicated impressionistically by the length of the bar.

3. 2 Paradigmatic Analysis

In the preceding section, some rules were specified for the realisation of the disfluency as either [X] [X] [X], or [X X X] or [X‾‾‾X] in which [X] is usually phonetically similar to part of the prevocalic consonants of the target syllable. In this section, the discussion will focus on a detailed evaluation of [X] itself and is applicable principally to [X‾‾‾X] disfluencies which were more common than the other two.

In the transcription system to be developed below, the convention is adopted that those attributes which, in auditory terms, are relatively invariant throughout the production of the disfluency are transcribed below, and at either end of, the bar. [m‾‾‾m] designates therefore that an [m] quality pervades the disfluency. Variations away from this constant auditory impression are transcribed as symbols, or diacritics, at various points above, or below, the bar.

3.2.1 Laryngeal Analysis.

The majority of disfluencies in the analysed corpus were either voiceless throughout, fully voiced or produced with a prolonged glottal stop: (5) - (7) show the transcriptions which have been used for these three laryngeal settings in the hypothetical disfluent production of a [w] initial syllable:

(5) w‾‾‾w (6) w‾‾‾w (with o o below) (7) w͡ʔ w͡ʔ

While it was possible to classify almost all of the disfluencies as either (5), (6) or (7) above, a detailed auditory and laryngographic analysis showed that either of the three phonatory settings could fluctuate to a different short-term setting requiring a transcription with a different symbol of diacritic. If this fluctuation is both audible and very short in duration, then a + symbol is used at a relevant point above the bar, the appropriate symbol or diacritic being transcribed below the + and bar. Thus the production correlate of (8)

(8) w‾‾+‾‾w (with ˷ below the +)

is a prolonged voiced labial-velar approximant with a sudden fluctuation to creaky voice at the point marked by the + symbol; furthermore, this fluctuation occurs about half-way into the prolongation, as indicated by the position of the + relative to the onset and offset of the prolonged continuant. By analogy, (9) - (11) are interpreted as follows:

(9) w‾‾+ +‾‾w (with o o below) (10) w‾‾+‾ʔ‾w (with o o below) (11) w‾‾+‾ʔ‾w

in (9)[1], the prolonged continuant is voiced, except for two short intervals of voicelessness at the points marked by +; in (10) and (11), the prolonged continuants are voiceless and voiced respectively except for a brief production of [w͡ʔ] at the points marked by the +.

In the transcription of stuttering disfluencies, it is often necessary to indicate changes in the phonatory setting which span an interval of time longer than the momentary change which corresponds to the + symbol. In this case, the interval is marked by two vertical lines on the bar, as in (12) and (13):

(12) n‾‾| |‾‾| |‾‾n (with o and ʔ below) (13) n‾‾| |‾‾‾‾| |‾‾n (with ˷ and o below)

In (12), the voiced alveolar nasal is produced continuously, except during two intervals: in the first, a voiceless nasal is produced (for the impressionistic interval indicated by the vertical lines); production of [n] subsequently continues until the next interval marked by the vertical lines in which [n͡ʔ] is produced; the remainder of the disfluency is produced as a fully voiced [n]. In (13), as in (12), a voiced alveolar nasal is produced continuously except for two intervals, one at the beginning (alveolar nasal with creaky voice) and one at the end (voiceless alveolar nasal). It is also possible to include + symbols within an interval of | |. In (14), for example,

(14) s‾‾|+ +|‾‾s (with z and ˷ ˷ below)

the disfluency is a prolonged voiceless alveolar fricative. For the interval between the vertical lines, the disfluency becomes voiced and, in addition, the onset and offset of the voiced interval are creaky, as indicated by the diacritics beneath the + symbols.

3.2.2 Supralaryngeal Analysis.

The transcriptions in (15) and (16) (target syllable *cap*) are used to describe two different kinds of repetition:

(15) [kʰ kʰap] (16) [k‾‾+‾‾kʰap] (with x below the +)

In (15), the first [kʰ] has the same phonetic characteristics as [kʰ] of *cap*. The transcription in (15) may correspond to what is habitually referred to in the literature as a stuttering *repetition*. The term 'stuttering repetition' is applicable to disfluencies of type (15) because a section of the prevocalic consonants is repeated once (or several times, usually at approximately equal intervals, as in Figure 2). In (16), the back of the tongue is raised to the velum forming a complete closure for an abnormally long duration. But in addition, this closure is punctuated by a release and turbulence at the point marked by the +. Apart from this momentary release, the closure forms a continuum, making the disfluency inseparable from the target syllable. It is quite possible that the release in (16) is involuntarily caused by the counteracting forces of high intra-oral air-pressure and an abnormally tense tongue-velum contact. If intra-oral air-pressure continually increases during complete occlusion, a point may come when the driving force of the former overcomes the resistance offered by the closure. In this case, the aerodynamic power may force a gap in the sealed tongue-velum contact, and this would result in a brief interval of frication.

It is often the case that the release and frication in disfluencies of type (16) are very short in duration. Figure 3, for example, shows releases which are so short that they are visible on spectrograms as a series of vertical lines that correspond to repeated burst onsets.

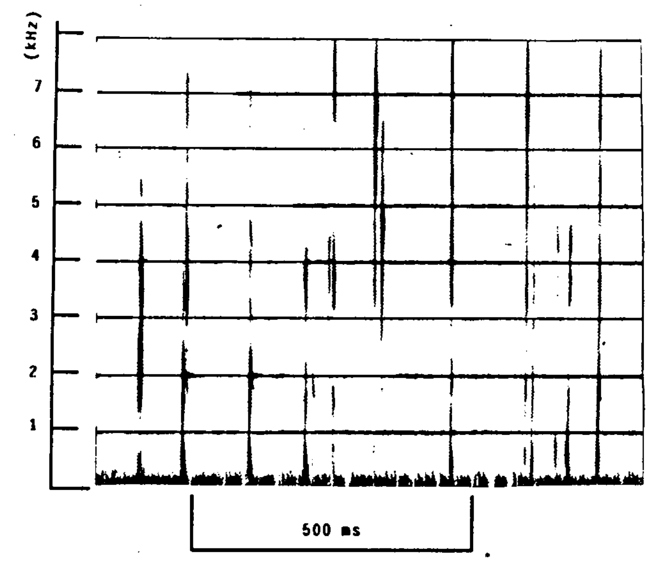

FIGURE 3: Spectrogram (300 Hz analysing filter) of a series of burst releases (target syllable *to*); two such releases occur at an interval of approximately 3 ms.

In order to distinguish these much shorter releases from those in (16), a single point is transcribed which corresponds to each release, as in (17):

(17) k‾‾+ +‾‾kʰap (with . . below)

Characteristic of many disfluencies is an increase in approximation between active and passive articulators probably caused by excessive tension throughout the vocal tract system, as suggested by Dalton & Hardcastle [6]. The closer approximation of approximants and fricatives resulting in a stop-like production is very noticeable in several subjects and the electropalatographic data often showed a considerable increase in the surface area of lingual-palatal contact in disfluently produced stops and affricates, as shown in Figure 4.

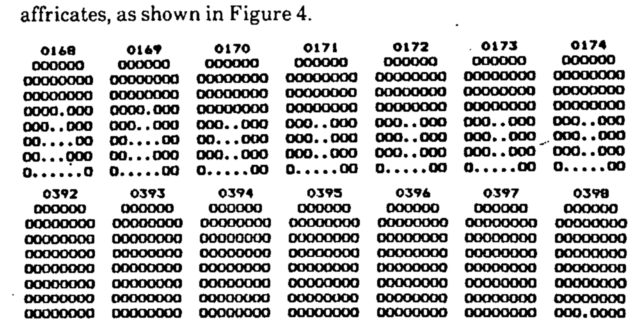

FIGURE 4: palatograms (sampling rate 100 Hz) of a prolongation of an affricate closure (target syllable *Jew*). Palatogram 168 occurs 210 ms after the onset of the closure; the increase in surface area of lingual-palatal contact (0 designates contact) is apparent from palatogram 392 (2.45 ms after the closure onset).

The progressively closer approximation of disfluently produced continuants can be transcribed in the following two ways:

(18) s‾‾st‾‾t (19) s‾‾t‾‾t

(18) corresponds to the production of a prolonged alveolar fricative which then became a prolonged alveolar stop at the location of the first [t] segment; in (19), the onset of the disfluency is an alveolar

fricative; thereafter, the degree of approximation gradually increases until a stop is produced. In (19), therefore, the closer approximation of the articulators is gradual, whereas in (18) it is comparatively abrupt.

Although the disfluency was most often realised as a section of the prevocalic consonant(s), occasionally there is a 'deflection' in the disfluency towards the vowel target. Thus, medially in a prolonged [n] continuant, for example, a vowel-like production might be audible which is very short in duration. A detailed acoustic and electropalatographic analysis of such vowel-like productions, reported in [7], has shown that they can consist of the entire acoustic vowel onglide of the corresponding fluently produced syllable, but not the acoustic vowel target. A spectrogram illustrating this phenomenon is shown in Figure 5 below.

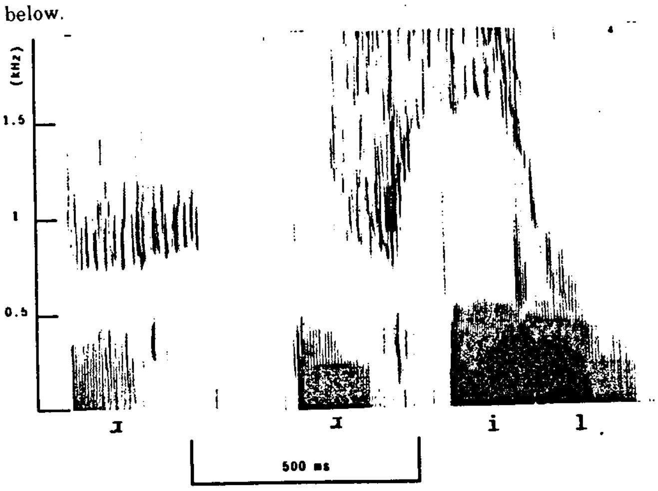

FIGURE 5: Spectrogram (300 Hz filter) of the offset of a prolonged [ɹ] continuant and the onset of *reel*. At the offset of the prolonged continuant, F2 rises less steeply to around 1 kHz compared with the F2 transition from [ɹ] to [i] in *reel* which rises to a target of 1.8 kHz.

The presence of the acoustic vowel onglide, or part of the acoustic vowel onglide, can be indicated by using the relevant vowel as a diacritic, as in (20):

+
e
(20) n n

4. CONCLUSIONS

An examination of a large corpus of stuttered speech [4] has shown that it is possible to make certain generalisations about the phonetic characteristics of disfluencies, at least when stutterers produce monosyllables in isolation; as far as possible, the design of the transcription system has been based on such general phonetic properties.

As discussed in more detail in [7], a stuttering disfluency can consist of any part of the syllable from its acoustic onset to the end of the acoustic vowel onglide (i.e., the acoustic vowel target and post-vocalic consonants are never realised as part of the disfluency). For the great majority of stutterers, part of this section of the syllable is prolonged in time. 'Prolonged in time' means that the dynamic change in shape of the supralaryngeal tract is minimal and that the disfluency is (in the majority of cases) either voiceless or realised as a glottal stop throughout its production. There are some, but less frequent, cases when the disfluency is fully voiced or produced with creaky voice from its onset to its offset. Disfluencies of type (2) and (3), in which an entire section of the prevocalic consonant(s) is repeated, are possible, but they are less frequent than type (4) disfluencies.

Characteristic of very many disfluencies seems to be a boost in the level of tension in the vocal tract system which may cause both the closer approximation of active and passive articulators (examples (18) and (19) and Figure 4) and some stop releases (examples (16), (17) and Figure 3). A more detailed empirical analysis is necessary to enrich this data.

Finally, it is possible, although comparatively rare, for part of the vowel to be realised during the production of a prolonged section of the prevocalic consonants (example 20); in such cases, the audible vowel-like sound is very brief and an acoustic analysis shows that the formant transitions 'bend towards', but never attain, the acoustic vowel target.

REFERENCES

[1] Wingate M.E. (1976) *Stuttering: Theory and Therapy*. Irvington: New York.

[2] Froeschels E. (1943) Survey of the early literature of stuttering, chiefly European. *Nervous Child* 2, 86-95.

[3] Morávek M. & Langová J. (1967) Problems of the development of the initial tonus in stuttering. *Folia Phoniatrica* 19, 109-116.

[4] Harrington J.M. (1986) *The Phonetic Analysis of Stuttering*. Ph.D dissertation, Department of Linguistics, University of Cambridge, England.

[5] Hardcastle W.J. (1972) The use of electropalatography in phonetic research. *Phonetica* 25, 197-215.

[6] Dalton P. & Hardcastle W.J. (1977) *Disorders of Fluency*. London: Arnold.

[7] Harrington J.M. (in press) An acoustic and electropalatographic study of stuttered speech. In Peters H. & Hulstijn W. (eds.) *Speech Motor Dynamics in Stuttering*. Springer Verlag: New York.

NOTES

1 For clarity, diacritics are transcribed a little further below the segment level than is usual to indicate a phonatory setting that is maintained *throughout* the disfluency (e.g. (10), in which the entire disfluency is voiceless, as opposed to (9), in which two brief intervals of voicelessness occurs medially in the disfluency).

2 See Terry M., Hiller S., Laver J., & Duncan G. (1986) The AUDLAB interactive speech analysis system. *IEE international conference on speech input/output: techniques and applications. Conference publication* 258, 263-265.