This study reports on the performance of a computerised digital signal processing system, known as the Computerised Frenchay Dysarthria Assessment (CFDA), which is designed to diagnose two sub-types of dysarthria – a family of speech disorders characterised by loss of control over the organs which facilitate speech production. This investigation explores the use of both spectral and waveform analysis to distinguish between the ataxic and mixed dysarthria subtypes by assessing articulatory competence in the execution of two speech-related tasks. It is demonstrated that waveform analysis of utterances representing the consonant /p/ can reliably measure a speaker’s lip seal competence; this combination of lip seal evaluation and voice quality measurement can then serve as a composite tool to detect certain pathological features characteristic of ataxic and mixed dysarthria respectively. To validate this hypothesis, this study compares the assessment accuracy of the CFDA application with that of a panel of expert clinicians when evaluating a series of speech samples from a selection of individuals, some of whom were previously diagnosed as suffering from either ataxic or mixed dysarthria. The CFDA’s diagnostic output from this data evaluation exercise produced an overall correlation of 0.91 with those of the expert clinicians. This close correlation reinforces the validity of the objective voice quality evaluation procedures developed during the course of this study.

Keywords: Dysarthria diagnosis, voice quality, digital signal processing, spectral analysis, F0 detection, lip seal

1. INTRODUCTION

The ability to produce well-inflected and clearly enunciated speech is dependent on the optimal functioning of the articulators, i.e., those organs which contribute to the process of oral communication. Any disruption in the functioning of the neuro-muscular mechanisms which control the articulators will usually cause degradation in the quality of speech, rendering it acoustically malformed and unusually difficult to decipher [1, 2]. The term dysarthria is used to describe those conditions that cause such neuro-muscular articulatory malfunction and the associated abnormal speech. Of course, a dysarthric condition can arise from impairment to one or more of the various speech subsystems, such as those which regulate respiratory or laryngeal functions. Given the vulnerability of each of these organ groups to disease or injury, the incidence of dysarthria – of either the permanent or temporary variety – is quite significant in the general population, estimated to be 170 per 100,000 in the United Kingdom [3]. Moreover, dysarthric symptoms are manifested in approximately a third of all individuals suffering traumatic brain injuries [4]. Dysarthric conditions will be manifested – to some degree – by at least nineteen percent of all those stricken with degenerative diseases such as multiple sclerosis (MS), Parkinson’s disease and motor neurone disease (MND). A similar frequency is noted among victims of stroke and other types of cerebro-vascular accident [5].

Given this relatively high incidence of dysarthria and the consequent need to further develop programs for appropriate rehabilitative treatment, one of the most pressing current research priorities is the design and implementation of objective diagnostic procedures to supplement existing subjective evaluation procedures which may give inconsistent results when used by inexperienced clinicians [6]. This study discusses a preliminary implementation of a computerised system of objective metrics, known as the Computerised Frenchay Dysarthria Assessment (CFDA), that is ultimately designed to distinguish five types of dysarthria via the use of digital signal processing (DSP) techniques incorporating spectral and waveform analysis of the speech signal.
The CFDA is itself derived from a paper-based diagnostic procedure – known as the Frenchay Dysarthria Assessment II (FDA 2), [7] – which is one of only two widely used dysarthria diagnostic testing protocols, the other being the Robertson Dysarthria Profile (RDP) [8].

As a continuation of a previous study by Carmichael [9], this report specifically focuses on evaluating the capacity of the CFDA to differentiate between ataxic and mixed1 dysarthria based on spectral and waveform analysis of digital audio recordings featuring dysarthric individuals attempting two speech-related tasks. It must be emphasised here that there appears to have been no prior attempt to create a system with the objective of DSP-based dysarthria-specific diagnostic functionality that is incorporated into the CFDA. Due to this lack of precedent, the CFDA’s diagnostic effectiveness will be validated by comparing the application’s diagnostic hypotheses with those of expert clinicians when both subjective (human) and objective (DSP) analytical techniques are applied to the same collection of digitised recordings featuring dysarthric speech samples.

The remaining five sections of this study are thematically arranged as follows: Section 2 reviews previous work in the domain while Section 3 outlines certain current diagnostic procedures used by the paper-based FDA. Section 4 discusses the CFDA’s functional requirements and the customised DSP algorithms it employs; Section 5 details various experiments conducted to validate the CFDA as a diagnostic tool and presents the results of these experiments; Section 6 proffers certain conclusions drawn from the experimental outcomes and Section 7 concludes the paper with recommendations for future research.

### 2. PRIOR WORK IN THE FIELD

As mentioned in the preceding section, it would appear that – although there have been several efforts to computerise some aspects of the paper based FDA’s functionality – the CFDA is the first computerised application that has implemented a range of DSP procedures capable of directly processing digitised speech data in order to produce a definitive diagnostic hypothesis specifying the type and severity of dysarthria, e.g. “patient appears to have mild ataxic dysarthria” or “patient appears to have moderate spastic dysarthria”2. Integral to the detection of dysarthria type is the analysis of the quality of phonation (voicing during speech) and the regulation of respiratory flow and pressure (i.e. the ability to control the flow of air from the lungs through the glottis and into the oral cavity). The following sections discuss previous investigative efforts concerning the measurement of these laryngeal functions.

1. Despite the all-inclusive nature that its name suggests, any specific case of mixed dysarthria usually incorporates only two dysarthria types, the most common combinations being Flaccid-Spastic and Ataxic-Spastic [7].
2. The Frenchay Dysarthria Test (FDT) – not to be confused with the paper-based FDA or the CFDA – is a software application developed by Roworth (as documented in Fitch [10]) which does not directly process speech data; the FDT processes only the FDA grades assigned by the presiding clinician when assessing a patient.

### 2.1 Spectral Analysis of Voice Quality

As reported by Carmichael [9], prior attempts to objectively evaluate voice quality have employed either speech signal spectral analysis techniques or some form of direct mechanical or visual inspection of the larynx during speech production. Videography techniques such as those used by Hsiao [11] – offer the most conclusive diagnostic evidence but are not practical for CFDA clinical examinations since most speech therapists lack the specialised training required to deploy the video-recording apparatus inside the patient’s laryngeal cavity. Mechanical inspection techniques are not as intrusive as visual inspection and typically require only the placement of some sort of sensory equipment on the patient’s neck. Unfortunately, the cost of such devices – known as electroglottographs or laryngographs [12, 13] – makes them an impractical proposition for the majority of small and medium-sized clinics. In contrast to these prohibitively expensive devices, software-based applications incorporating DSP-based spectral analysis techniques are a relatively low-cost option more easily acquired by public health-care institutions. The following paragraphs review the most widely-used spectral analysis techniques currently available.

Spectral analysis of voice quality, which consists in the description of phonation texture in terms of component audio frequencies, is a well researched discipline boasting more than four decades of investigation. Given the scope of this current inquiry, this literature review will focus on prior research in voice quality measurement which attempts to categorise instances of pathological voice according to Enderby’s classification [7]. In the context of dysarthria diagnosis, Enderby recognises three main types of abnormal voice quality: guttural, strained/strangled and breathy [7]. As detailed in Table 1, the manifestation of these aberrant voice types can help to differentiate between the five main dysarthria sub-groups (these sub-types of dysarthria being ataxic, Extrapyramidal, flaccid, mixed and spastic [7]). To date, the objective description/evaluation of breathy voice has proven a straightforward task in the spectral domain due to the elevated levels of high-frequency noise (i.e. breathiness) caused by the incomplete shutting of the vocal cords during the closed phase of the glottal cycle while voicing [14]. It would appear that the most reliable method of identifying breathiness in phonation is by comparing the levels of energy in the lower and upper frequency bands of the speech signal. Both Shoji et al [15] and Li et al [16] have used variations of high frequency/low frequency energy ratio to consistently distinguish between breathy and normal voice samples. As discussed in a subsequent section of this report, the CFDA incorporates a measurement algorithm employing a variation of this energy ratio measure in order to determine the severity of breathiness for a given voice sample.

<table>
<thead>
<tr>
<th>Dysarthria Type</th>
<th>Associated Voice Pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxic</td>
<td>Guttural</td>
</tr>
<tr>
<td>Extrapyramidal</td>
<td>Strangled/strangled or breathy</td>
</tr>
<tr>
<td>Flaccid</td>
<td>Breathy</td>
</tr>
<tr>
<td>Mixed</td>
<td>Strained and/or breathy, occasionally guttural</td>
</tr>
<tr>
<td>Spastic</td>
<td>Strangled/strained, harsh</td>
</tr>
</tbody>
</table>

1. Despite the all-inclusive nature that its name suggests, any specific case of mixed dysarthria usually incorporates only two dysarthria types, the most common combinations being Flaccid-Spastic and Ataxic-Spastic [7].
2. The Frenchay Dysarthria Test (FDT) – not to be confused with the paper-based FDA or the CFDA – is a software application developed by Roworth (as documented in Fitch [10]) which does not directly process speech data; the FDT processes only the FDA grades assigned by the presiding clinician when assessing a patient.
The HNR may be defined as: the maximum autocorrelation, from Hiraoka et al [21] where they dub is greater than zero. Using a variation of HNR analysis which Hr percent of the speakers in the pathological voice group exhibited an analysis [20]. In essence, HNR measures better precision has been obtained using the harmonic-to-noise energy ratio (SNR) and Normalised Noise Energy (NNE) measures – such as those employed by Emanuel & Sansone [19]; however, better precision has been obtained using the harmonic-to-noise energy ratio (HNR) analysis [20]. In essence, HNR measurements represent the difference in energy intensities between the harmonic and non-harmonic components of the speech signal. The HNR may be defined as:

\[ HNR(\text{dB}) = 10 \times \log_{10} \frac{r_2'(\tau_{\text{max}})}{1 - r_1'(\tau_{\text{max}})} \]  

from Boersma [20] where the degree of periodicity across the frequency ranges of a signal, \( x(t) \), is calculated by computing the maximum autocorrelation, \( r_i'(\tau_{\text{max}}) \) at the time lag \( \tau \) which is greater than zero. Using a variation of HNR analysis which they dub relative harmonic intensity analysis (\( H_r \)), Hiraoka et al [21] report an accuracy rate of 95% in identifying of hoarse and normal voices for a selected group of speakers with normal and pathological voices. Essentially, Hiraoka’s \( H_r \) metric is a ratio of the sum of the energy intensities of the second and higher harmonics expressed as a percentage of the overall harmonic intensity, including that of the fundamental frequency. The relative harmonic frequency energy ratio may be defined as:

\[ H_r = \left( \frac{\sum_{i \geq 2} P_i}{P} \right) \times 100(\%) \]  

from Hiraoka et al [21] where \( P_i \) is the spectral energy at the \( i^{th} \) harmonic and \( P \) is the total energy including that contained in the fundamental frequency. Hiraoka observed that ninety per cent of the speakers in the pathological voice group exhibited an \( H_r \) of less than 67% (i.e. the harmonics contributed less than two thirds of the overall energy) while 94% of speakers from the non-pathological group manifested \( H_r \) values greater than this 67% threshold value. Despite their exceptional results, the classification criteria of Hiraoka et al. seem somewhat inadequate in one important aspect: their categorisation of the abnormal features of pathological voice is not sufficiently discriminatory: Hiraoka divides the speakers into just two categories, namely hoarse and normal, and thus implicitly suggests that “hoarseness” should be considered a generic category encompassing all types of abnormal voice texture. Such a broad generalisation of voice pathology is inadequate for the diagnostic objectives of this CFDA prototype which – for the purpose of distinguishing ataxic from mixed dysarthria – requires classification functionality capable of designating a given phonation sample as either hoarse or guttural or breathy. Section 4 of this study discusses the application of the \( H_r \) metric for CFDA voice quality evaluation. In the following section, however, we focus our attention on various techniques used to measure respiratory pressure.

### 2.2 Analysis of Respiratory Pressure

In the context of dysarthria diagnostics, respiratory flow and pressure are normally assessed either as an isolated function or in the context of some speech activity. For non-speech respiratory pressure evaluation, current dysarthria evaluation protocols – such as the FDA and RDP – normally require the patient to demonstrate competence in regulating respiratory flow by attempting a controlled prolonged exhalation; in contrast, respiratory pressure is usually assessed by observing the patient’s ability to produce some form of plosive burst (i.e. an explosive expulsion of air from the vocal tract, typical in the production of stop consonants such as /t/ or /k/). In the case of the FDA, the patient is instructed to produce a series of ten bilabial plosive bursts – in the form of repeated utterances of the phone /p/ – so as to demonstrate competence in producing good lip seal (i.e. the capacity to initiate an increase in air pressure within the oral cavity and then suddenly release the pressurized air in the form of a bilabial plosive burst which would normally produce /p/).

The quality of lip seal is usually assessed audiovisually and/or by placing a hand directly in front the patient’s lips to gauge the force of the air emission generated by the plosive burst. Alternatively, some type of computerised strain gauge – such as the Entran Flatline pressure transducer [22] – may be used to measure respiratory pressure at the point of lip seal. Such systems, however, require expensive proprietary hardware components, rendering them unsuitable to be incorporated into the CFDA.

The following section discusses the development of bespoke CFDA digital signal processing algorithms to facilitate the task of correctly classifying phonation and lip seal acoustic data as manifesting symptoms of either ataxic or mixed dysarthria.

### 3. THE CFDA DSP DIAGNOSTIC CHALLENGE

As mentioned previously, the classification by Enderby [2, 3, 7] specifies five main dysarthria sub-types and their associated voice pathologies (Table 1); it is to be noted that ataxic dysarthric individuals are almost invariably guttural when phonating while those who suffer from the mixed variety of dysarthria tend to display a breathy voice quality. A principal hypothesis of this investigation is that these differences in voice abnormality be-
The substantial dissimilarities between the atactic and mixed groups in terms of voice quality and lip seal competence have made it possible to develop a prototype version of the CFDA which classifies digitised audio recordings (of sustained /a/ phonations and plosive /p/ utterances) as belonging to one of four categories, i.e.: [i] ataxic dysarthria [ii] mixed dysarthria [iii] other dysarthria sub-type or [iv] Normal (i.e. showing no signs of acoustic abnormality). It is to be noted that this four-way diagnostic system represents only an initial proof-of-concept implementation to demonstrate the feasibility of using DSP techniques in this domain; the ultimate objective, however, is to produce a CFDA application that completely emulates all the functionality of the paper-based FDA (which provides guidelines to aid in the diagnosis of symptoms from any of the five dysarthria sub-types). It is also noteworthy that the paper-based FDA consists of twenty-six individual tests, each of which incorporates evaluation guidelines for the purpose of assigning a numerical grade for a given execution of some FDA test item; these numerical grades range from a minimum of '1' (for the poorest quality efforts) to a maximum of '9', a score indicating that no abnormality has been detected. By way of illustration, the evaluation criteria for the lip seal and phonation tasks are presented in Tables 2 and 3.

Table 2 FDA Subjective Assessment Criteria for Sustained /a/ Phonation Test [7].

<table>
<thead>
<tr>
<th>FDA Grade [Max=9]</th>
<th>Performance Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 or 9</td>
<td>Patient can say &quot;AH&quot; clearly for 15 seconds.</td>
</tr>
<tr>
<td>6 or 7</td>
<td>Patient can say &quot;AH&quot; clearly for 10 seconds.</td>
</tr>
<tr>
<td>4 or 5</td>
<td>Patient can say &quot;AH&quot; for 5 to 9 seconds, but phonation may be interrupted by intermittent huskiness or breaks in phonation.</td>
</tr>
<tr>
<td>2 or 3</td>
<td>Patient can say &quot;AH&quot; for 3 to 5 seconds clearly.</td>
</tr>
<tr>
<td>1</td>
<td>Patient unable to maintain a clear phonation on &quot;AH&quot; for 3 seconds. Voice continually strained/strangled or gutteral.</td>
</tr>
</tbody>
</table>

Table 3 FDA Subjective Assessment Criteria for Lip Seal Test.

<table>
<thead>
<tr>
<th>FDA Grade</th>
<th>Performance Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 or 9</td>
<td>Repeats &quot;/p/ /p/&quot; with even seal</td>
</tr>
<tr>
<td>6 or 7</td>
<td>Occasional air leakage, break in lip seal, or lip seal not consistent for plosion on each sound</td>
</tr>
<tr>
<td>4 or 5</td>
<td>Lip seal observed on sound, but auditorily weak.</td>
</tr>
<tr>
<td>2 or 3</td>
<td>Very poor lip seal, pressure lost from one segment of the lips. Patient able to attempt closure but unable to maintain. Not auditorily represented</td>
</tr>
<tr>
<td>1</td>
<td>Patient unable to maintain any pressure. Patient unable to visually or auditorily represent sound.</td>
</tr>
</tbody>
</table>

After a patient completes an FDA diagnostic test session, the test administrator plots the resulting grades on an x-y axis to produce what is known as an FDA bar chart. It is normally the case that the juxtaposition of these bar charts will generate certain contours or profiles that correspond to specific sub-types of dysarthria. Figure 2 shows a complete FDA bar chart profile typical of an individual suffering from ataxic dysarthria.

Figure 1 presents a series of dual-element bar charts depicting typical /p/ lip seal and /a/ phonation gradings for five individuals (one from each of the five dysarthria sub-groups); all of these persons were diagnosed as manifesting a moderate level of dysarthric affliction.

The substantial dissimilarities between the mixed and ataxic groups can serve as reliable diagnostic indicators.

In addition to the abovementioned voice quality differences, Enderby’s 1986 comparative study [23] also observed that those suffering from mixed dysarthria manifested – on average – the poorest lip seal performances among the five groups (see Figure 1).

Figure 1 Dual element bar chart profiles for 5 individuals from each of the Five Dysarthria sub-groups.
In addition to using objective measurement techniques to evaluate speech data, the CFDA application in its current implementation is also designed to assign an FDA-style numerical grade to evaluate any attempt at steady-state /a/ phonation or repeated /p/utterances. The diagnostic challenge for the CFDA, therefore, is to generate FDA numerical grades which are identical or very similar to those given by expert clinicians when assessing the same digitised audio recordings of dysarthric speech data. Such a human-machine comparison is the inspiration for the experimental procedures described in section 5 of this study.

The CFDA’s sequence of diagnostic operations when evaluating speech data is schematically represented in Figure 3.

4. CFDA DSP ANALYSIS

As mentioned previously in Section 1, the CFDA is designed to operate as a diagnostic tool in real-world conditions using only low-cost hardware resources; typically, the sole hardware component required during the course of a CFDA diagnostic session is a standard head-mounted microphone, such as those used for internet telephony applications. The CFDA’s spectral analysis and waveform measurement techniques for the evaluation of voice quality and lip seal respectively will now be discussed in greater detail.

4.1 Quantifying the Force of Plosive Burst

Given the stated preference for employing low-cost software-based DSP technology in the CFDA application, the two viable options for assessing plosive burst would be some form of spectral evaluation or a technique using time-domain analysis (an example of such as a time-domain analytical technique is described subsequently). After examining the spectral footprint of a series of /p/utterances of both acceptable quality and otherwise, it was observed that the waveform profile of a plosive burst proved to be the most consistent indicator in terms of identifying successful attempts at producing a well-formed plosive burst. The paragraph that follows discusses the procedures for extracting a waveform representation of an audio signal.

During the process of digitisation, an acoustic signal is sampled (measured) at regular intervals and each of these samples is assigned a positive or negative numerical value. The plotting of these sampled values on an x-y axis produces a waveform representation of the digitised signal – with time and signal intensity measured along the x and y axes respectively. The sudden increase in acoustic energy levels occasioned by a well-formed plosive burst (see Figure 4) causes a steep peak in the waveform, particularly if the microphone recording the signal is placed directly in front (i.e. within three centimetres) of the speaker’s mouth.

A cross-comparison of the spectral data derived from recordings made by normal and dysarthric speakers performing the /p/ lip seal task has revealed certain distinctive patterns which are characteristic of a well-formed bilabial plosive:

The onset of the /p/ plosive burst produces a spike in the signal waveform, the steepness of which is in direct proportion to the intensity of the plosive burst. Well-formed bilabial plosions usually occasion a waveform spike manifesting a gradient of at least 3:1.

The duration of the plosive burst is no longer than ten milliseconds and is usually followed immediately by vowel onset, producing an utterance which may be phonetically transcribed as /pe/ and is acoustically similar to the initial “puh” consonant-vowel sequence in the English word “put”.

It has proven possible – using the abovementioned features as classification criteria – to detect instances of malformed lip seal in the context of a CFDA diagnostic evaluation. It has been

---

4 These criteria for well-formed bilabial plosives were established in consultation with a group of speech and language therapists who listened to a selection of recordings featuring /p/ utterances produced by normal and dysarthric individuals; the therapists then identified individual instances of /p/ utterances which were deemed of acceptable quality.
observed that lip seal attempts which are malformed/defective usually exhibit one or more of the following pathological phenomena:-

- **Inability to maintain respiratory pressure**: Persons with poor lip seal competence usually exhibit unintentional air leakage from the lips just before executing the plosive burst; such leakage usually causes any attempt at /p/ production to be perceived as some form of fricative, such as /f/.

- **Irregularity in /p/ execution**: In some cases, even if one or more well formed /p/ utterances are produced, it will be evident that the speaker experiences some difficulty in producing a series of ten /p/ utterances in quick succession. As a result, the intervals between utterances are often irregular and of varying intensity (i.e. the first attempts are substantially more energetic compared with the last two or three attempts).

- **Weak plosive burst**: as mentioned previously, well-formed lip seals normally generate a waveform spike with a gradient of at least 3:1; /p/ utterances not exhibiting such a gradient are considered to indicate some degree of incompetence at regulating respiratory pressure.

Given the abovementioned threshold values which differentiate lip seal quality, it has proven possible to devise a system of metrics to objectively classify a series of lip seal attempts realised as /p/ utterances. In essence, these measurement units quantify the differences in maximum amplitude for each of a given series of plosive bursts and the consistency of the inter-plosive intervals. In the waveform representation of the lip seal attempts shown in Figure 5, for example, the intervals between the ten plosive attempts range from a tenth of a second to as much as 2.91 seconds while the sampled values corresponding to the amplitude peaks vary between 13739 and 6473.

These variations in amplitude maxima and spacing can be expressed in terms of the mean average difference in plosive peak values and the standard deviation in the duration of the inter-plosive intervals respectively.

In the case of the lip seal attempts depicted in Figure 5, the measurements would be calculated as follows:

(i) Lip seal attempts 3, 5, 7 and 9 do not satisfy the gradient threshold requirements to be classified as acceptable plosives, thus only the other six attempts will be considered.

(ii) The consecutive spacing (measured in seconds) between the six acceptable plosives are as follows: 0.10, 2.11, 2.42, 2.91, and 2.81. The standard deviation derived from these values corresponds to 1.15. This measurement of the regularity of inter-plosive spacing will henceforth be referred to as the inter-plosive spacing variation (IPSV). It has been noted that IPSV values greater than 0.7 correspond to a subjective impressions of excessive variability.

(iii) The amplitude peaks for the considered plosive peaks are the following: 13739, 8039, 7516, 7141, 6073 and 7919. The difference in peak values can be expressed as the ratio comparing a given peak to its immediate successor. A ratio of 1:1 would indicate that the maximum value for a given peak is the same as that of the one immediately following; conversely, a 1:0.50 ratio means that a given peak’s maximum value is double that of the next in the series. In the lip seal series depicted in Figure 5, the ratios between the adjacent peaks of interest are as follows:- 1:0.59, 1:0.93, 1:0.93, 1:0.95 and 1:1.31. Consultation with the expert clinician group has confirmed that ratio differentials in excess of ±30% (i.e. ratios greater than 1:1.30 or less than 1:0.70) indicate unacceptable inconsistency for a contiguous pair of plosive bursts. Such inconsistency in lip seal production can be expressed as a plosive deviation index (PDI), which is computed as the mean average of all instances of deviation exceeding the aforementioned 30% differential threshold. In the case of the lip seal series depicted in Figure 2, two of the ratios – 1:0.59 and 1:1.31 – exceed the threshold by eleven percent and one per cent respectively, resulting in a PDI score of 12.

In addition to the objective descriptors to measure lip seal competence, the CFDA also incorporates specialised voice quality spectral analysis techniques, the specifics of which are detailed in the next section.

### 4.2 CFDA Spectral Classification of Voice

Given the application’s diagnostic objectives in terms of differentiating between ataxic and mixed dysarthria, it is necessary for this preliminary implementation of the CFDA application to recognise at least three types of voice pathology, namely (i) breathy (ii) hoarse and (iii) guttural. For the measurement/identification of breathy voice, the high-frequency energy distribution (HED) index developed by Carmichael [9] is used again be for this investigation. In essence, the HED calculates – for a given steady state vowel phonation – the percentage of the signal’s energy present in the 6KHz-8KHz frequency band; as Carmichael [9] has demonstrated, steady state /a/ phonations

---

5In order to ascertain the IPSV threshold corresponding to perceptions of irregularity in /p/ sequencing, a SUFDAR recording was selected which featured 10 plosive bursts of acceptable quality and with an IPSV of 0.2. This recording was then manipulated (by way of shortening or lengthening inter-plosive segments of silence) so as to produce 10 versions of the original recording, with each version manifesting a different IPSV value (ranging from 0.1 to 1.0). The expert group of therapists recruited for the various experiments in this study was invited to assess all ten of the recordings to determine which of them featured a /p/ sequence with an unacceptably irregular rhythm. The recordings with IPSV values of 0.7 and above were unanimously identified as excessively irregular.
containing more than fifteen percent of such high frequency energy are usually considered by expert clinicians to be inordinately breathy. Accordingly, an HED score indicates—in terms of percentage points—the amount of high frequency energy in an audio signal which exceeds the 15% threshold. For example, a sustained /a/ phonation manifesting 20% of its energy in the 6 KHz–8 KHz frequency band would be assigned an HED of 5 (indicative of the 5% excess over the 15% threshold). According to Carmichael [9], HED values ranging from 1 to 4 were considered by expert assessors to be slightly breathy while HED scores greater than 4 but less than 7 were deemed moderately breathy; /a/ utterances manifesting an HED of 7 or higher were unanimously perceived as very breathy.

For the assessment of hoarse and guttural voice, the CFDA application uses a combination of measures, including Hiraoka’s $H_r$ ratio and a pitch jitter measurement developed by Carmichael [9]. For the CFDA application, the $H_r$ ratio is expressed as a value indicating the contribution (as a percentage) of the second and higher harmonics to a given signal’s overall energy. The reader is reminded that if the energy contribution of the higher harmonics is less than 67% of the total energy in a vocalisation, then said vocalisation is normally perceived as being hoarse. Furthermore, it has also been observed that the level of pitch jitter can be used to differentiate between hoarse and guttural voice qualities; this phenomenon will be discussed in greater detail in the following paragraphs.

![Pitch Contour for Mod-erately Guttural /a/ Phonation](image)

**Figure 6** Screen Shot of CFDA Phonation test showing pitch contour for moderately guttural /a/ phonation.

As detailed by Carmichael [9], pitch jitter is manifested in any unintended / erratic changes in pitch (also known as glottal frequency or F0 frequency) during the course of a phonation. It must be noted that—when processing an audio signal—the CFDA generates an F0 value for every 20-millisecond segment of the signal. Rapid oscillations in glottal frequency (as is the case with the pitch contour for the /a/ vocalisation in Figure 6) normally indicate hoarse and/or guttural voice quality.

Given the presumption that glottal frequency should remain relatively constant during a steady state vowel production, pitch jitter may be defined as the difference in pitch values between two contiguous 20-millisecond segments. Accordingly, it is possible to express pitch jitter as the standard deviation for a series of values representing the inter-plosive percentage difference in frequency for adjacent 20-millisecond segments (provided that the F0 values for such contiguous segments is greater than zero6). In order to compensate for any differences in utterance length (and the resultant greater likelihood of a higher standard deviation for longer vocalisations), a variant of the standard deviation algorithm is used, this variant, known as the population standard deviation, may be expressed as follows:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x - \bar{x})^2}$$

from Everitt & Dunn [24] where a value of one is subtracted from the divisor $N$ (which is the sample size) in order to compensate for the aforementioned likelihood of larger variation for bigger sample sizes [24]. This population standard deviation for the observed F0 values will henceforth be referred to as the pitch variation index (PVI).

Carmichael [9] demonstrated that PVI may be used to classify hoarse voice texture; during the course of this investigation, it was established that the PVI measure can also serve as an instrument for the detection and evaluation of guttural phonation since the two voice pathology types share certain spectral characteristics, i.e. both hoarse and guttural voice types are characterised by an elevated PVI and $H_r$ count. It is possible, however, to distinguish between hoarse and guttural voices based on the range of F0 frequencies observed in the phonation signal: according to the subjective evaluations of expert speech therapists, guttural

---

6Since it is occasionally the case that the pitch extraction algorithm may fail to locate the pitch peak for a given frame and therefore return an erroneous value of zero for said frame, such zero-value frames—provided that they do not occur consecutively in a sequence exceeding four in number—are not considered when calculating the pitch variation index (PVI).
voices – in comparison to hoarse voices – exhibit greater pitch jitter but lower F0 minima: the expert clinicians considered as guttural all but one of the phonation samples which exhibited a modal average pitch frequency of less than 160 Hz and a PVI exceeding 0.50. Given the foregoing, PVI and $H_r$ measurements are used to objectively describe and classify both hoarse and guttural voice. Table 4 details the mapping of HED, PVI, PDI and ISPV values to the various FDA numerical grades.

The following section discusses the experimental procedures used to validate the CFDA’s dysarthria-type classification protocols based on the evaluation of lip seal and phonation.

5. VALIDATING CFDA DIAGNOSIS

The principal hypothesis of this investigation is that it is possible to differentiate between the ataxic and mixed dysarthria subtypes based solely on an assessment of a speaker’s voice quality and competence in maintaining lip seal. Moreover, it is intended that the CFDA’s analysis of acoustic data and resultant classification should emulate the diagnostic accuracy of expert clinicians when conducting “on-the-spot” assessments of a patient’s oral responses to CFDA test stimuli. In order, therefore, to make possible such a comparison, a group of three expert clinicians – one of whom is the author of the original paper-based version of the FDA – was invited to grade two collections of recordings7 featuring a variety of normal and dysarthric speakers attempting to produce steady state /a/ phonations and – as a separate task – a series of ten /p/ utterances.

These recordings are drawn from a larger corpus of digitised dysarthric speech data known as the Sheffield University Frenchay Dysarthria Assessment Recordings (SUFDAR). For this experiment, ten individuals were selected from each of the five dysarthria sub-groups; the dysarthric typology and degree of affliction of these fifty individuals are given in Table 5. Apart from the dysarthric sufferers, ten normal persons (i.e. not suffering from any dysarthric condition) were also used as a control group.

<table>
<thead>
<tr>
<th>Dysarthria Type</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxic</td>
<td>3</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Extrapyramidal</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Flaccid</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Mixed</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Spastic</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The expert panel assessed the phonation and lip seal attempts according to the criteria provided by the paper-based FDA (see Tables 2 and 3); as discussed previously, these evaluation guidelines grade a performance using a numerical grade, ranging from one to nine points. In addition to assigning a numerical grade to each performance, the members of the expert panel were also requested to identify the voice texture of each phonation attempt as belonging to one of the following categories: (i) breathy, (ii) hoarse, (iii) guttural and (iv) normal.

In order to calibrate the CFDA’s speech processing algorithms to enable them to output an FDA-style numerical grade as well as a voice pathology or lip seal competence classification, a series of digitised phonation and lip seal recordings (which are not part of the SUFDAR corpus) were graded by the expert panel; the members of said panel were then interviewed by the CFDA software developers in order to determine the threshold values distinguishing a test performance worthy of an FDA grade “9” as opposed to a performance only meriting, for example, a grade “7”. The subjective assessments of these clinicians were then compared to the objective DSP measurements of the CFDA to establish the correspondence between a given FDA grade and a specific range of computerised objective measurements. As mentioned previously, the CFDA’s objective grade allocation criteria are detailed in Table 4.

Tables 6 and 7 present a selection of the subjective (human) and objective (machine) grading assessments for two dysarthric individuals attempting the /a/ phonation and lip seal tasks. A more detailed discussion of this comparison between objective and subjective assessment is the subject of the next section.

5.1 Experimental Results

As shown in Table 8, the CFDA achieved a rate of classification accuracy corresponding to 88.3%, correctly identifying the

Table 6 Expert Clinician and Automatic CFDA Evaluations for 2 Dysarthric Speakers Attempting /a/ Phonation Task.

<table>
<thead>
<tr>
<th>Speaker ID</th>
<th>Grade</th>
<th>Evaluator ID</th>
<th>Observations/ Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxic #061</td>
<td>8</td>
<td>001</td>
<td>Slightly breathy</td>
</tr>
<tr>
<td>8</td>
<td>002</td>
<td>Mildly breathy</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>003</td>
<td>Mildly breathy</td>
<td></td>
</tr>
<tr>
<td>CFDA</td>
<td>PVI = 0.08; HED = 1.2; $H_r$ = 66.4%. Segment of voicing which is clear = 12.5 s (out of 15.8 s).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed #031</td>
<td>3</td>
<td>001</td>
<td>Moderately guttural</td>
</tr>
<tr>
<td>5</td>
<td>002</td>
<td>Slightly guttural</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>003</td>
<td>Moderately guttural</td>
<td></td>
</tr>
<tr>
<td>CFDA</td>
<td>PVI = 0.41; HED = 0.33; $H_r$ = 73.0%. Segment of voicing which is clear = 0 s (out of 6.3 s).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7 Expert Clinician and Automatic CFDA Evaluations for 2 Dysarthric speakers attempting the /p/ Lip Seal Task.

<table>
<thead>
<tr>
<th>Speaker ID</th>
<th>Grade</th>
<th>Evaluator ID</th>
<th>Observations/Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxic #061</td>
<td>9</td>
<td>001</td>
<td>Good lip seal demonstrated</td>
</tr>
<tr>
<td>9</td>
<td>002</td>
<td>/p/ auditorily strong</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>003</td>
<td>Clean lip seal</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CFDA</td>
<td>10 acceptable plosives; IPSV = 0.36; PDI = 0.87</td>
<td></td>
</tr>
<tr>
<td>Mixed #031</td>
<td>5</td>
<td>001</td>
<td>Air leakage noted, weak seal</td>
</tr>
<tr>
<td>3</td>
<td>002</td>
<td>/p/ auditorily weak, inconsistent</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>003</td>
<td>/p/ produced as /l/, irregular spacing</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CFDA</td>
<td>4 acceptable plosives; IPSV = 0.36; PDI = 0.87</td>
<td></td>
</tr>
</tbody>
</table>

Table 8 CFDA Classification Performance for SUFDAR data.

<table>
<thead>
<tr>
<th>Correctly Classified</th>
<th>Ataxic</th>
<th>Ataxic classified as Other</th>
<th>Mixed</th>
<th>Mixed classified as Other</th>
<th>Other</th>
<th>Other classified as Ataxic</th>
<th>Other classified as Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

dysarthria sub-type (or lack thereof) manifested by fifty-three of the sixty persons whose recordings were submitted for evaluation. It is noteworthy that the CFDA correctly diagnosed eighty percent of both the ataxic and mixed cases; furthermore, there was only one instance where the CFDA confused a case of mixed dysarthria with the ataxic variety. In that particular example, the individual’s pathological symptoms were more characteristic of ataxic than mixed dysarthria since his voice quality was adjudged as guttural (by both the expert clinicians and the CFDA). Moreover, this individual’s performance of the lip seal task evidenced extraordinary lip seal competence for someone suffering from mixed dysarthria, a variety of the disease which is usually characterised by a substantially reduced ability to control respiratory pressure within the oral cavity.

In terms of the CFDA’s capacity to emulate the diagnostic skill of expert clinicians when evaluating specific recordings featuring attempts at sustained /a/ phonation and /p/ lip seal, an overall correlation of 0.91 was computed for the numerical grades returned by the CFDA and those of the expert clinicians as a group. In order to calculate this human-computer correlation, the CFDA’s scores were compared to the mean average of the scores given by the clinicians for each of the SUFDAR recordings used in the various test corpora for the experiments conducted in this study. In the case of the lip seal attempt by participant #031 (see Table 7), the scores accorded by the three clinicians (‘5’, ‘3’ and ‘5’) produce an average score of 4.33, which is then compared with the CFDA score of ‘4’ to calculate a correlation co-efficient. Repeating this process for all the human/machine scores for all the data evaluated produces a correlation coefficient of 0.91, indicating that the CFDA’s diagnostic decision-making closely emulated that of the expert clinicians who participated in this study. It is also noteworthy that the CFDA’s correlation with the clinicians as a group was stronger than the scoring correlation between expert clinicians #001 and #002, whose scores only yielded a 0.88 correlation. A more comprehensive interpretation of these experimental results is provided in Section 6.

6. CONCLUSIONS

A combination of both customised and well-established spectral measurement techniques have been successfully employed to objectively diagnose three types of voice quality pathology, namely guttural, hoarse and breathy voice textures. Of the sixty phonation samples evaluated, there were only two instances where the voice quality description returned by the expert clinicians differed substantially from that of the CFDA application. In addition, waveform gradient analysis has been used to reliably identify /p/ utterances which demonstrate satisfactory lip seal competence as contrasted with /p/ utterances which show evidence of poor lip seal. It must be noted, however, that there were 8 cases (out of 60) where the CFDA’s lip seal evaluation differed materially from that of the expert clinicians, but this still represents an 87.67% level of accuracy, which the clinicians themselves deemed acceptable.

The generally successful implementation of the aforementioned objective evaluation protocols has the cumulative effect of demonstrating the feasibility of using acoustic DSP analysis for differentiating between ataxic and mixed dysarthria sub-types. The CFDA’s 80% classification accuracy is satisfactory given that this performance is based on data from only two of the twenty-six tests needed by the paper-based FDA to formulate a diagnostic hypothesis. Moreover, the CFDA’s data visualisation techniques and implementation as a stand-alone application makes it a viable and practical real-world tool for practising clinicians.

After further development, the CFDA application is scheduled to undergo field trials with a select group of speech and language therapists from the British National Health Service. If these field trials prove successful, the CFDA may well become as widely used as its paper-based counterpart, the FDA 2.

7. RECOMMENDATIONS FOR FUTURE RESEARCH

Given this study’s successful proof-of-concept implementation demonstrating the feasibility of employing DSP techniques to distinguish instances of ataxic from mixed dysarthria, it is the intention of this research team to computeraise all of the other twenty-four FDA tests so as to enable the CFDA application to identify the typology and severity of any dysarthric condition in accordance with the five dysarthria sub-types proposed by Enderby. Of course, the complete automation of the assessment protocols for the other FDA acoustic tasks will require the further elaboration of novel DSP algorithms to measure a variety of other articulatory abnormalities – such as hypernasality – which have hitherto not been the focus of any large-scale speech technology investigation. It is with great enthusiasm, therefore, that the
CFDA developers strive to make a further contribution to this area of computerised speech measurement and classification.

ACKNOWLEDGMENTS

This researcher acknowledges the assistance and contribution of the University of Sheffield, UK, and the AlGhurair University, UAE, in supporting this investigative effort and its immediate predecessor. This researcher also takes this opportunity to pay tribute to his father, Nathaniel Carmichael, whose work as a forensic analyst has inspired his son to also attempt some contribution to the domain of biometric classification.

REFERENCES