Jeffery, T. and Cunningham, S. and Whiteside, S. P. (2017) Analyses of Sustained Vowels in Down Syndrome (DS): A Case Study Using Spectrograms and Perturbation Data to Investigate Voice Quality in Four Adults With DS. The Journal of Voice. ISSN 0892-1997

This is an Accepted Manuscript published by Elsevier in its final form on September 21, 2017 at http://dx.doi.org/10.1016/j.jvoice.2017.08.004.

Analyses of sustained vowels in Down Syndrome (DS): a case study using spectrograms and perturbation data to investigate voice quality in 4 adults with DS.

Objectives. Automatic acoustic measures of voice quality in people with Down Syndrome (DS) do not reliably reflect perceived voice qualities. This study used acoustic data and visual spectral data to investigate the relationship between perceived voice qualities and acoustic measures.

Study design. Participants were 4 young adults (2 male, 2 female; mean age 23;8 years) with DS and Severe Learning Disabilities (SLD), at least one of whom had a hearing impairment (HI).

Methods. Participants imitated sustained /i/, /u/ and /a/ vowels at pre-determined target pitches within their vocal range. Medial portions of vowels were analysed, using Praat, for Fundamental frequency (f_0), HNR, jitter and shimmer. Spectrograms were used to identify the presence and the duration of subharmonics at onset and offset, and mid-vowel. The presence of diplophonia was assessed by auditory evaluation.

Results. Perturbation data were highest for /a/ vowels and lowest for /u/ vowels. Intermittent productions of subharmonics were evident in spectrograms, some of which coincided with perceived diplophonia. The incidence, location, duration and intensity of subharmonics differed between the four participants.

Conclusions. Although the acoustic data do not clearly indicate atypical phonation, diplophonia and subharmonics reflect nonmodal phonation. The findings suggest that these may contribute to different perceived voice qualities in the study group and that these qualities may result from intermittent involvement of supraglottal structures. Further research is required to confirm the findings in the wider DS population, and to assess the relationships between voice quality, vowel type and physiological measures.

Key Words: Down Syndrome, diplophonia, subharmonics, voice, ventricular, phonation

INTRODUCTION

Individuals with DS are described as having voices that are characteristic of the syndrome [1, 2] with descriptors commonly identifying ‘harsh’, ‘guttural’, and ‘raucous’ qualities [3, 4, 5, 6]. Additionally, DS voice is often perceived as being atypically breathy and rough [2, 7] and low-pitched [2, 3, 7]. Several studies have investigated voice qualities in adults and children with DS, using automatic measurements of fundamental frequency (f_0), jitter (frequency perturbation), shimmer (amplitude variation)) and harmonic-to-noise ratio (additional noise in the harmonic spectrum) [2, 5, 8, 9, 10, 11, 12, 13]. Historically, these measures have been used as indicators of atypical phonation, although there is increasing evidence that they are unreliable for pathological voices [14]. Research using such measures has failed to determine the role of phonation in perceived voice qualities in DS populations. Inconclusive and contradictory findings from acoustic measures have led some researchers to suggest that supralaryngeal factors, such as the properties of the vocal tract, may contribute to the characteristic voice qualities more than laryngeal factors [10, 15].

Several studies have compared acoustic data to the perceived vocal qualities of people with DS, with conflicting results [9, 10, 11]. Lee et al. [10] analysed the vowels from continuous speech of nine British adults with DS, aged between 17-24 years, and those of typically-developing (TD) controls, matched for age and sex. They reported no clear difference between groups in jitter and shimmer values. Using data derived from spoken words, Albertini et al. [11] reported higher mean f_0 and lower spectral energy in DS adults, compared to controls, and reduced shimmer in adult males with DS. In contrast to both studies, Moura et al. [9] reported elevated measures of jitter and shimmer in the sustained vowels (/a/, /e/, /i/, /o/, /u/) of 66 Portuguese children with DS, aged 4-8 years. In comparison to data collected from TD children, the DS children produced sustained vowels at a lower f_0, with greater deviation, higher perturbations in shimmer and jitter, and with increased noise in the signal in comparison to the strength of the phonation. The DS children performed statistically differently on all voice measures except for the f_0 of the vowel /u/. In comparison to TD controls, the children with DS were found to have lower measures in spectral tilt (ST) [9], a measure of the energy across the frequency ranges. Spectral tilt can indicate creakiness (strong positive slope) or breathiness (strong negative slope) [16]. The authors suggested the finding indicated higher than typical levels of breathiness and more ‘forceful’ phonation.

Few studies have examined sustained vowel production in adult DS populations. Of these, there is agreement that mean f_0 is high, compared to controls [2, 8, 13], but findings are mixed and are difficult to reconcile with perceived qualities. Moran and Gilbert [2] compared acoustic data to auditory-perceptual judgements in 16 adults with DS. They reported elevated jitter (<6%) in three of their participants and additional noise in the harmonic spectrum of nine participants. Despite an atypically high mean f_0 in the DS group, low pitch was perceived by 70% of the judges in 5 of the participants. For females, mean f_0 correlated with perceived pitch, but there was low correlation between perceived pitch and f_0 for DS males. The authors suggest that the perceived voice quality in DS males is affected by the interaction of several factors, including HNR, breathiness and laryngeal tension. A mismatch between perceived voice quality and acoustic data was also reported in a more recent study [13]. Seifpahani et al. [13] analysed the sustained /a/ vowels of 22 adults with DS. Jitter was lower than controls, but shimmer was comparable. Their data did not indicate perturbation, despite agreement from three speech and language therapists (SaLTs) that all participants were 'moderately hoarse'.

An early study by Beckman et al. used spectrographs to examine the sustained that were produced by a female (ages 22;8 years) and a male (27;6) with DS [8]. Both participants had normal hearing levels and voices that were described as breathy, with an imbalance in oral/nasal resonance. They identified that in six of the nine vowels produced by the female subsequent cycles of voicing were more variable in duration (jitter) and in amplitude (shimmer). They reported that regular alternations in periods of the waveform resulted in an effective halving of f_0 and a perceived octave drop in pitch. The female was subsequently identified as having diplophonia, which is the generation of two audible pitches [17, 18, 19]. Beckman et al. [8] suggested that sub-glottal variations in pressure or laryngeal pathology might have contributed to the phenomenon in their subject.

Beckman et al. [8] suggested that a high incidence of diplophonia in the DS population might explain reports of a lower perceived pitch. A decade earlier, Novak [4] had proposed that ventricular voice, which is caused by the continued oscillation of the ventricular folds, is the cause of the perceptually harsh voice quality in DS subjects. Beckman et al. [8] argued that hypotonia ruled out the probable engagement of the ventricular folds in their subject. However, it is now known that people with DS apply more energy than controls with healthy voices to trigger contraction in the surface of the laryngeal musculature [15]. The habitual use of excessive effort in producing voice can result in hyperfunctional voice disorders [20, 21] in which supraglottal structures, such as the ventricular folds, are employed during phonation [17, 22, 23, 24]. One recent study with children and young people with DS used auditory-perceptual evaluation to identify laryngeal tension and diplophonia [12]. Diplophonia was not perceived, although elevated levels of laryngeal tension were perceived. However, diplophonia is not reliably perceived by auditory evaluation alone [18].

Recent studies of DS voice rely on automatic acoustic measures, and have not included data derived from visual inspection of individual samples. Perturbation analysis is unreliable for aperiodic voice [17]. Therefore, visual inspection may be necessary to confirm whether samples are valid for automatic analysis, and to provide additional information regarding the possible presence of diplophonia [18]. The current study examines the acoustic characteristics of phonation in sustained vowels produced by four adults with DS, SLD and HI. The study uses voice perturbation data (Harmonics-to-Noise Ratio [HNR], jitter and shimmer) and mean fundamental frequency [f_0] alongside evidence from visual data to explore the nature of phonation in sustained vowels, and to consider how the data link to the participants' habitual voice qualities. Although sustained vowels are not a reliable indicator for voice quality in speech [25] the use of sustained vowels allows for examination of phonation when articulation and processing demands are low. As such, it can be expected that any difficulties that are revealed in these conditions will be exacerbated in connected speech [17, 25, 26].

2. METHODS

2.1 Participants

The study was a multiple case-study design that involved four participants. Data were collected from four young adults (mean age: 23;8 years; SD=0.37) with DS and SLD, who were participating in a larger, explanatory study that examined voice production in speech and song. Explanatory case studies seek to explain causal links between phenomena that cannot be understood through experimental studies [27]. They are appropriate for investigating under-researched or poorly understood aspects of behaviour, such as voice production in DS, and heterogeneous groups, such as those with DS. Data were generated for each participant separately. Participants acted as their own controls.

2.1.1 Demographic data

As part of the larger study, the group had participated in a range of standard and non-standard cognitive tasks. This established that the individuals in the group were of similar abilities in verbal mental age (MA), as measured by the British Picture Vocabulary Scale, version 1 (Tables 1-4). The BPVS test [28] measures receptive vocabulary, and correlates to verbal intelligence. Participants are required to listen to a word and identify the corresponding picture, from a set of four. It has been standardised for use for children and adults to 17;11 years, and has been successfully used with people with DS to estimate MA across a range of ability levels. Information about voice quality and hearing ability was provided by the participants' SaLT. Descriptors of voice quality for each participant were also given by the SaLT, based on her existing knowledge and historical evaluations of each participant. In order to establish vocal range, participants were asked to imitate an ascending and descending vocal glide on /a/ vowel. To assess speaking range and mean f_0, participants were asked to describe a picture [29] and to provide positive comments on a peer's performance of a song.

2.1.2 Recruitment

Ethics approval was granted by the Human Communication Sciences' Ethics Committee, University of Sheffield. Participants gave informed consent to take part in the study and for their data to be used.

All participants were resident in long-term care and informed consent was sought and obtained from the residential care home in the first instance. The Principal of the care home identified possible participants using the inclusion criteria, which were:

1. a diagnosis of DS;

2. aged between 11-25 years;

4. a recognised degree of speech impairment; and

5. an interest in singing.

Pictorial information sheets were provided for potential participants that explained the research aims and methods in outline. Detailed written information sheets were provided for staff and parents/carers. Staff within the organisation distributed letters to parents/carers and to potential participants, together with information sheets. The same staff sought consent from the participants after a period of two weeks. Four young people were identified and approached, and all four gave informed consent. All participants had the right to refuse to participate, or to withdraw; consent was therefore ongoing, and not all participants completed all tasks.

2.2 Stimuli

The vowels /a/, /i/ and /u/ were demonstrated to each participant at pitches that were within their vocal range, as determined by their vocal range (see Tables 1-4). Target notes were C, G or E, and target pitches were: l31 Hz (C3), 165 (E3), 196 Hz (G3), 262 Hz (C4), or 330 Hz (E4), or 392 Hz (G4). As part of the wider study of musical ability a range of pitches was used to assess accuracy in pitch matching across the vocal range. Performance of these vowels at high mid and low pitches has been used to measure HNR, jitter and shimmer using Praat, in male and female patients with healthy voices and with dysphonia [30]. For healthy voices there is no statistically significant difference as a result of pitch in perturbation measures (jitter, shimmer, HNR), but a significant pitch effect has been noted in voices with dysphonia [30].

In the present study, the first author played each pitch on a chime bar, then imitated it vocally at a pitch deemed to be within the participant's vocal range. Participants were given the instruction: 'listen to the sound, then sing the same sound for as long as you can'. As the primary aim was to measure pitching accuracy, if the participant struggled to reproduce the note at the pitch given, it was presented again at the same pitch. If the participant still had difficulty in matching the given pitches of G4 or C4, these were presented at an octave lower (196 Hz - G3); 131 Hz - C3), if these were deemed more appropriate to their habitual vocal range. If necessary, demonstrations were also repeated to encourage improved vowel imitation, or a lengthier production. Therefore, the number of imitations, and the number of vowels at specific target pitches, differed between participants. All productions by the participants were retained for analysis, but some were later discarded if they were contaminated by noise (see Section 2.4).

2.3 Recording procedures

Audio recordings were made using a shock-mounted cardioid condenser RODE NT1A microphone that was set to ‘unidirectional’ mode. The microphone was connected to an iMac via an M-Audio Mobile-Pre USB soundcard, and recorded onto the laptop using Garageband 08 Version 4.1.2 (Apple inc. 2002-2007) at a sampling rate of 44.1 kHz. A pop-shield was placed 8 centimetres from the front surface of the microphone to encourage participants to maintain a constant distance. The recording levels were adjusted as necessary using the ‘input’ dial on the soundcard, located to the researcher’s right, to maintain a constant level. It was not possible to control for levels of sound pressure level (SPL) or for lung volume.

The assessment was conducted in a room familiar to the participants during the second week of a six-week programme of group singing tuition. Participants were assessed individually, but all were present for the task, which formed part of formative assessment of singing abilities. There was occasional noise pollution from the adjoining room, which affected the quality of some recordings. In addition, some participants joined in with the assessments of others, leading to further contamination. These issues were addressed during analysis (see Section 2.4).

2.4 Preparation of vowels for acoustic and visual analysis.

Recordings for each participant were inspected visually and aurally in Praat, vs. 5.4.05 [31]. Samples that were contaminated by continuous noise or by noise during the medial portion of the vowel were excluded entirely from further analysis. This resulted in 25% of recorded samples being rejected, and fewer valid samples being available for some participants. Contamination also occurred in some samples if the onset of the participant’s vowel overlapped with the offset of the demonstration. Where this occurred, the overlap was noted on a text file in Praat [31], and the contaminated section was excluded from acoustic analysis. Individual whole vowels were isolated from the recordings and saved as .wav files. Each file was imported separately into Praat, a program that has been used in previous studies with people with DS [9, 10, 11, 12].

Medial portions of vowel were selected for acoustic measures. Schaeffler et al. [26] report that the time taken for voices to stabilise in connected speech and sustained vowels is longer for disordered voices, and can exceed 70 ms. Accordingly, the initial and final 100 ms were discarded from perturbation analysis. Sections of vowel were selected manually. The full vowel was selected first, with reference to a broadband spectrogram. The onset was determined by the point at which the pitch contour began. Offset was determined by the end of the pitch contour or the end of a stable first formant contour. The visual cues were then used in Praat to segment the vowel, using TextGrid, enabling select and generate data for each vowel section.

2.4.1 Calculation of subharmonics.

For all valid vowels, narrowband spectrograms were generated in Praat in order to determine the incidence and duration of multiple harmonics between dominant harmonics, and single subharmonics. Visual analysis can provide objective measures of phonation [32] and is especially informative where voicing is atypical [33, 34]. This method enabled detailed analysis of voice production in those samples where a high degree of aperiodicity prevented automatic measurements of voice perturbation (see Figure 1 for an example of aperiodic voicing and Figures 1-4 for examples of main harmonics and subharmonics). Studies of sustained vowels typically exclude onset and offset [25] which can result in the exclusion of data that may be particularly relevant to disturbed phonation [26, 35, 36]. Therefore, three measures of subharmonics were generated for each sustained vowel: the initial 100 ms (onset); the final 100 ms (offset); and the medial portion between the onset and offset (T-200 ms).

Settings for narrowband spectrograms were based on Cavalli and Hirson [18]: window length was set at 15 ms, the frequency range was 0-1000 Hz, and the dynamic range was adjusted to 35 dB to screen out the effects of background noise. Text grids were used in Praat to demarcate and annotate sections of the vowel. Sections of atypical phonation as indicated by the presence of subharmonics on the narrow-band spectrogram were highlighted. Their duration was noted in the text grids and their duration was expressed in milliseconds according to their position (initial, medial, offset), and as a percentage of the total measurable vowel duration

(T- contaminated sections). The presence of diplophonia was confirmed in sections containing subharmonics with reference to the audio recording for that section in isolation.

2.4.2 Relative intensity of subharmonics

Finally, for each participant, images of spectrograms were generated for their /a/ vowel that contained the highest percentage of subharmonics (Tables 1-4). The /a/ vowel was chosen for illustration, as it is a common stimulus in studies involving participants DS [8, 9, 10]. Images were generated at a dynamic range of 35 dB and 15 dB. Images at 35 dB and at 15 dB allowed visual comparison between participants of the relative intensity of subharmonics, which may be indicative of pathology [34]. The technique was used to provide greater understanding of how spectral data might relate to perceived voice quality; and to consider how similarities and differences in spectral data between participants might reflect known differences in habitual voice quality.

2.5 Auditory-perceptual judgement

The description of voice types for each participant was provided by the participants' SaLT, based on historical records.

Judgement was made by the first researcher as to the presence or absence of diplophonia in sustained vowels. This process was informed by the visual data and auditory evaluation. The presence of subharmonics is associated with diplophonia, but also with related perceptual qualities such as creak, pitch breaks, or roughness [18, 19]. However, diplophonia can also occur in the absence of subharmonics [18]. Therefore, sections of sustained vowel that contained subharmonics were listened to through headphones, and were compared to the sections of the same vowel that did not contain subharmonics. If subharmonics were present for the vowel's duration, the vowel was compared to another sample by the same participant, that did not contain subharmonics. A judgement was made by the first researcher as to the presence or absence of diplophonia. Although perceptual evaluation of diplophonia is also unreliable [18], previous studies have relied upon judgement of whole vowel samples. This study enabled repeated listening to sections within vowels, and comparison within or across samples. The aim was to determine whether visual acoustic data could be a useful tool in understanding or distinguishing voice types in the participants.

2.6 Acoustic Analyses

Recordings were imported as .wav files into Praat for automatic measurement of voice perturbation measures. Measures of mean f_0, HNR, jitter (%), and shimmer (%) were generated. These measures are commonly reported in voice research in DS literature [9, 10, 11, 12]. The standard settings were used for all measurements and for the report template. The cross-correlation method was used to calculate pitch within Praat. The pitch range was set to 50-400 Hz in order to capture the low-frequency vocal productions.

Data were coded as non-diplophonic or diplophonic, based on information from spectrograms in conjunction with auditory evaluation by the first author (see Section 2.5).

2.7 Intra-measurer Reliability

After a period of at least 12 months, the following data from were re-measured by the first author, from the original recordings:

- Voice perturbation measures generated by Praat for the first two valid productions of each participant's sustained /a/, /i/ and /u/ vowels (67%); and
- The duration and percentage of subharmonics in sustained vowels, at onset and offset, and in medial sections. Five sustained vowels were retested for three participants (M1: 55%; F2: 55%) and F1: 45%); and four were retested for the fourth participant (M2: 44%), who produced fewer vowels at onset that were suitable for measures.

An interclass correlation (ICC) test was applied in SPSS v21. The results indicated high intra-measurer reliability, with statistically significant ($p<0.05$*) or highly-significant ($p<0.01$**) effects. For one participant (M1), the data for the percentage of subharmonics were less reliable (ICC = 0.379; p= .530). On examination of the data, this result was influenced by the initial inaccurate measurement of the duration of subharmonics at offset in one vowel only, of the order of 12 ms.

3. RESULTS

3.1 Case Study 1 (M1)

Male 1 (M1) has unconfirmed hearing loss (Table 1). His SaLT record states that his hearing is 'adequate for speech' but that he has difficulties discriminating low frequency sounds, affecting his ability to perceive whispered speech. According to his SaLT, his voice is loud, gruff and harsh in quality. His vocal range on pitch glides spanned 13.59 semitones, and his range in the speech tasks exceeded 18 semitones. However, in speech his lower range was produced within vocal fry range [37].

INSERT TABLE 1

Of his nine vowels, jitter exceeded published norms for six; shimmer and HNR were close to the published means for all vowels but /a/1 [30]. All vowels were produced with subharmonics at onset and offset. Only two vowels (/i/1 and /u/3) were produced without subharmonics medially. The vowel /a/1 was produced with continuous subharmonics (100%) For the remaining eight vowels, the percentage duration of subharmonics was below 27%. Diplophonia was perceived in one vowel (/a/1: Table 1).

The vowel (/a/1) is shown in Figure 1. Figure 1a shows multiple harmonics from onset to offset, which fluctuate in intensity. The dominant harmonics occur at multiples of about 96 Hz, and subharmonics at about 48 Hz. Diplophonia was perceived as continuous. The vowel was produced with very low HNR, and high jitter and shimmer (Table 1). At 15 dB, the subharmonics remain visible in the region of 500-1000 Hz (Figure 1b).

INSERT FIGURE 1

3.2 Case Study 2 (M2)

M2 has confirmed hearing loss (Table 2). A recent audiogram by the college SaLT showed a loss in his right ear of 50-60 dB between 250-1000 Hz, and at 4000 Hz; and a loss of 46 dB at 2000 Hz. The loss in his left ear is 40 dB at 250 Hz, and between 60 - 65 dB for the range 500-4000 Hz. M2 did not wear a hearing aid: pictures or Makaton were used during assessments to support his comprehension of task. His voice quality was described by the SaLT as ‘quiet, breathy and pubophonic’. His pitch range spanned 16.45 semitones but when speaking his vocal range was limited to less than 4 semitones. His mean f_0 in the speaking tasks were close to his minimum pitch for vocal glides. Although the vowels were presented at pitches that were within M2’s vocal range as measured by glides, the majority of these were close to or above the upper f_0 of his speaking range (Table 2).

INSERT TABLE 2

Of his nine valid vowels, jitter exceeded published norms for one vowel (/a1/) but was within norms for four vowels (/a/3, and /u/ 1, /u/ 2, and /u/3); shimmer exceeded norms in two vowels (/a/ 1 and /a/2), and was within norms for four vowels (/i/1, /i/3, /u/1, /u/ 2, and /u/3); HNR was within norms for six vowels. His productions of six vowels overlapped with the end of the demonstration, affecting measurement of subharmonics at onset, and for parts of medial segments of some vowels. Subharmonics were present medially in 6 of 8 valid vowels, and at onset of all three valid vowels. No subharmonics were present at offset. The percentage duration of subharmonics was 22.67% or less. Diplophonia was perceived in two vowels (/a/1 and /a/2).

The vowel /a/1 is shown in Figure 2. The subharmonics overlapped at onset with the end of the demonstration, but re-emerged towards the end of the vowel. The figure shows two subharmonics at different frequencies between the main harmonics. Three sections were produced with audible diplophonia, each lasting approximately 125 ms in duration; these coincide with the presence of single subharmonics.

INSERT FIGURE 2

3.3 Case Study 3 (F1)

F1 is described by her SaLT as having an unspecified degree of bilateral hearing loss, but no difficulties in perceiving speech sounds (Table 3). Her voice was described as harsh and rough, and sometimes appeared low-pitched. Her vocal ranges in glides and when describing pictures were 6.61 and 6.11 semitones, respectively, and her vocal range was 8.94 semitones in spontaneous speech. Jitter was within norms for all four /i/ vowels and for /a/3; shimmer was within norms for all /u/ vowels, and for three of the four /a/ vowels; HNR was within norms for seven vowels (all /u/ vowels, /i/2 , /a/3, and /a/4) [30].

INSERT TABLE 3

Subharmonics were present in all valid vowels: no data were available at onset or mid-vowel for /u/3 as a result of overlap with the demonstration. Subharmonics were evident in seven of eight valid vowels at onset, 8/9 medially (/a/4) and in four of seven vowels at offset. Diplophonia was perceived in five vowels (/a/1, /a/3, /i3/, /i/4, /u/2: subharmonics exceeded 62% of the vowel duration in all diplophonic vowels.

Figure 3 shows the vowel /a/1, which was produced with low HNR and high jitter (Table 1). The waveform shows single that were present at onset and re-emerged mid-vowel. These coincided with audible diplophonia.

INSERT FIGURE 3

3.4 Case Study 4 (F2)

F2 has a mild but untested hearing loss (Table 4). Her voice range spanned over an octave when imitating pitch glides (19.58 semitones), and exceeded 17 semitones in both speech tasks. Values for jitter, shimmer and HNR were within the range of published norms [30] for all vowels except /a/1 and /a/4. Subharmonics were evident at onset for all valid vowels (n=8); and occurred in medial position for seven vowels. Diplophonia was perceived in one of these vowels (/a/4), in which the percentage duration of SH was 81.13%.

INSERT TABLE 4

Figure 4 shows multiple interharmonics at onset, which reduced to a single subharmonic without a clear break (above the second main harmonic: Figure 3). Diplophonia was perceived in segments containing single subharmonics. No subharmonics were visible at 15 dB (Figure 4b).

INSERT FIGURE 4

4. DISCUSSION

4.1 Perturbation measures and perceived voice qualities

The perceived habitual voice qualities of participants are partially reflected in the perturbation measures. Although contested [14], elevated jitter may be consistent with a ‘rough’ or ‘harsh’ voice quality, and shimmer may be associated with ‘breathy’ qualities [19]. Jitter was elevated in most tokens produced by M1 (67%) and F1 (55%), both of whom have ‘harsh’ sounding voices (Table 1 and Table 2, respectively), and in two tokens by F2 (22%), whose voice is less consistently harsh (Table 3). However, jitter was also elevated for M2 (55%), whose voice is not described as rough or harsh (Table 2). Shimmer values were elevated in five vowels for F2 (45%) and four vowels (44%) in M2, both of whom have typically breathy voices. Shimmer was elevated in two vowels in F2 (22%), and one vowel for M1 (11%). HNR was within norms in over 67% of productions for all participants.

The data indicate inconsistent productions of the same vowel, and differences according to vowel quality. Across the group, the vowel /a/ resulted in the greatest incidence and degree of perturbation for individuals, and /u/ the least (Tables 1-4). This is consistent with data from healthy voices that demonstrate a greater tendency for English-speaking adults to produce creaky voice qualities in low vowels (/a/, /ae/) than in high vowels (/i/, /u/)/ [38]. A distinction between vowel quality and voice quality was noted by Moura et al. in their study of Portuguese children with DS [9]. The authors proposed that a reduced forward movement of the tongue in low vowels reduces laryngeal tension and lowers the fundamental frequency, affecting stability of phonation and resonant properties [9].

The study supports previous findings that the atypical voice qualities described of adults with DS is not necessarily evident in acoustic voice data [9, 10, 11, 12]. However, it is well known that sustained vowels are poor indicators of voice quality in speech [39, 40]. It is therefore unsurprising that the perceived voice qualities in conversational speech that differentiate M1 and F1 from F2 and M2 (Table 1) may not be reflected in the mean voice data alone. Furthermore, subharmonics create 'noise' in the signal that make it impossible to measure features that depend upon a periodic signal [41, 42, 43]. Although jitter and shimmer have been linked to perceptions of roughness and breathiness, respectively [19], several studies have concluded that they are unreliable indicators of voice quality [41, 44, 45, 46], especially for voices perceived to score highly on measures of Grade, a measure of perceived hoarseness [14]. Furthermore, jitter and shimmer can present as normal in voices that contain subharmonics [47] and are unreliable for diplophonic voices [48, 49]. For all participants in the present study, this is evident in those vowels that contain subharmonics but result in normative data (e.g. M1: /a/3; M2; /u/2 and /u/3; F1: /u/1; F2: /u/1), including vowels perceived as diplophonic (F1: /a/3, /u/2) Therefore, it is questionable as to how reliable and valid it is to use measures that rely on periodicity of voicing for a population whose voice is consistent with measures of dysphonia, such as hoarseness, breathiness, roughness and strain.

4.2 Perceived qualities and visual data

Evidence from visual inspection of all vowels (Tables 1-4) suggest that each participant has difficulties in initiating and sustaining modal phonation, and that two (M1 and F1) had difficulties in terminating phonation. The duration of subharmonics at the onset of vowels exceeded 100 ms for most participants, which is indicative of atypical or 'disordered' voices [26]. Difficulties in initiating and terminating phonation will have a considerable impact on phonation in speech [17, 39]. It is therefore possible that these portions of vowels may be more useful indicators for speech difficulties than medial sections.

One possible explanation for the atypical duration of subharmonics at onset is that the participants engage their false vocal folds (FVF) more intensely or for longer than is typical when initiating phonation. Although the FVF are activated at the onset on phonation in healthy voices [22, 24], their dynamics at onset can affect the stability and intensity of phonation [22, 50]. The visual data from Figures 1-4 might be indicative of the involvement of the FVF or other supraglottal structures. The subharmonics that are observed in the sustained vowels of F1 (Figure 2), F2 (Figure 3) and M2 (Figure 4) are lower in intensity than the main harmonics. Their productions are perceived as diplophonic which can arise from oscillation of the FVF [18, 19]. Comparison of the visual data for M1 (Figure 1) with published a published source [51] suggest that M1 might engage additional supraglottal structures. His /a/ vowel is diplophonic, but his spectrogram is complex, compared to Figures 2-4: there are multiple harmonics, which modulate in intensity and overlap, resulting in no clearly defined main harmonic. The spectrogram of his pathological vowel is comparable to those produced in 'growl' phonation, in which the aryepiglottic fold vibrates [52]. Sakakibara et al. [51] report that the arypeiglottic folds vibrate at half the rate of the FVF, generating additional harmonics. The spectrogram that Sakakibara et al. used to illustrate growl phonation shows two strong fundamentals in the frequency range below 2 kHz, a peak at about 1.5 kHz, and weak fundamentals above this range. Their image is comparable to that of M1's /a/ vowel (Figure 1).

Differences in perceived voice qualities within the group (Tables 1-4) are reflected in the relative intensity of subharmonics in the spectrograms of sustained /a/ vowels. The relatively intense subharmonics evident in the productions of M1 and F1 (Figures 1 and 2, respectively) are consistent with studies that link the presence of subharmonics to ‘harsh’ [18, 19, 48] and ‘creaky’ [52] vocal descriptors. Although such studies also report an inexact correlation between acoustic data and auditory-perceptual evaluation, this approach may be a promising adjunct for understanding phonation, using non-invasive means.

4.3 General Discussion

Diplophonia was evident in all participants in one or more samples. Rodger [12] found that diplophonia was not present in the voices of 22 children and young adults with DS. However, diplophonia is difficult to perceive in trained listeners and is often confused with ‘creaky’, ‘harsh’ or ‘rough’ qualities [18, 19, 48], which were reported in Rodger’s [12] study. It remains possible that diplophonia is present in the voices of people with DS, but that it may not be apparent in either perceptual judgements or in mean perturbation data. This study suggests that intermittent diplophonia and a high degree of subharmonics may contribute to the typical descriptors of voice qualities for people with DS. Furthermore, it is possible that the presence of subharmonics gives the impression of harsh, rough or low-pitched voices within the DS population. Kramer, Linder and Schönweiler [34] reported a correlation between high subharmonic content and their raters’ perceptions of low f_0 in 145 speakers with dysphonia and rough voices. They also found that the intensity of subharmonics, and the percentage of low f_0 values were also linked to perceived roughness. They conclude that the percentage of subharmonics in speech increases the perception of low-pitched noise and contributes to the perception of roughness.

Previous studies have discussed the potential involvement of FVF in DS voice [4, 8]. Beckman et al. [8] and Rodger [12] dismissed the involvement of FVF as a contributory factor to diplophonia in people with DS based on lax vocal muscles. However, Braunschweig et al. [53] argue that in TD populations, elevated levels of pathology at onset can indicate hyperfunctional disorder that may result from weak musculature. Despite reports of laryngeal hypotonicity in DS, excessive laryngeal tension at onset of phonation has been measured in adults [15]. Abnormal tension within the laryngeal muscles or in the external muscles can result in a harsh’ or ‘strained’ quality [20, 21, 24], ventricular dysphonia [20, 24] and diplophonia [18, 19, 33, 54]. It is possible that excessive tension may account for the incidence of intermittent diplophonia in all four participants in the present study, and may distinguish the habitual voice qualities of M1 and F1 (table 1, Table 3). In this study, the greater intensity of vibration in M1 and F1's subharmonics (Figures 1 and 2, respectively) suggests a greater involvement of the supralaryngeal structures [22, 24].

Although the data confirm intermittent diplophonia and atypical perturbations that are consistent with hyperfunctional voice disorders [42], physical examination would be needed to confirm the role of muscular tension in these participants. Information regarding hearing ability would also be required to confirm the impact of hearing loss. HI is associated with hyperfunctional voice disorders in HI populations [56], as well as with pubophonic voice and diplophonia [55, 56, 57]. It is probable that HI contributed to M2's voice production in these tasks. This information was unavailable for M1, F1 and F2. Given that HI affects up to two thirds of people with DS [58], it is possible that the unspecified degrees of hearing loss (Table 1) could contribute to voice difficulties in all participants in this study.

5. Conclusions

The findings indicate that all four participants intermittently produce atypical phonation, but that this does not necessarily result in elevated perturbation measures. As with previous studies [25, 35, 41], this study questions the validity of using mean acoustic measures (jitter, shimmer, HNR) that are derived from steady-state vowels to shed light on voice quality within the DS population. The use of visual data alongside auditory information was instrumental in understanding the limitations of the acoustic data, and for shedding light on the nature of voice production in the participants. The number and intensity of subharmonics might prove useful indicators for voice quality in those with DS, specifically at onset and offset. It is also suggested that harsh voice qualities may be further differentiated through visual inspection of spectrograms. However, this study is based on a small sample size, and upon unequal samples from each participant, both in terms of number of valid vowels, and duration of vowels. To verify the findings, further large-scale research would be needed within this population, with greater consistency between samples, and with reference to control groups. Future studies could examine further the prevalence of diplophonia and the link between the number and intensity of subharmonics and perceptual correlates. A number of automatic acoustic measures have recently been developed that might be usefully applied for large-scale research, such as Sun's Subharmonic-to-harmonic ratio [59] or the recently developed Diplophonia Diagram [48].

The results are consistent with an intermittent difficulty in initiating, in maintaining or in terminating phonation. They suggest that a physiological difficulty contributes to atypical voice quality in the speech of adults with Down Syndrome speech. Specifically, the visual data provide support for Novak's proposal [4] that ventricular production may account for the characteristic harsh, low-pitched voice quality in people with DS. However, this study was unable to assess levels of hearing, which could result in similar voice difficulties, or to monitor physiological responses during vowel production. Such measures would be necessary to confirm the link between perceived voice quality, acoustic measures and physiological factors.

Acknowledgements

The authors would like to thank the participants of the study.

References

[1] Moran, MJ. Identification of down's syndrome adults from prolonged vowel samples. *J Commun Disord*. 1986; 19(5):387-394. doi:10.1016/0021-9924(86)90028-6.
[2] Moran, MJ, Gilbert, HR. Selected acoustic characteristics and listener judgments of the voice of Down syndrome adults. *Am J Ment Defic.* 1982; 86(5):553-556
[3] Benda, CE. Mongolism. In CH Carter (Ed.), Medical Aspects of Mental Retardation. 1965; Springfield: Illinois
[4] Novak, A. The voice of children with Down's syndrome. *Folia Phoniatrica* (Basel). 1972; 24:182–194.
[5] Pentz Jr AL, Gilbert HR. Relation of selected acoustical parameters and perceptual ratings to voice quality of Down syndrome children. *Am J Ment Defic*. 1983; (2), 203-210.
[6] Montague jr, JC, Hollien, H, Hollien, PA, Wold, DC. Perceived pitch and fundamental frequency comparisons of institutionalized Down's syndrome children. *Folia Phoniatr et Logop*. 1978: 30(4), 245-256.
[7] Montague JC, Hollien, H. Perceived voice quality disorders in Down's syndrome children *J Commun Disord*. 1973; (6), 76-87
[8] Beckman, DA, Wold, DC, Montague, JC. A noninvasive acoustic method using frequency perturbations and computer-generated vocal-tract shapes. *J Speech Hear Res*. 1983;26(2):304-314.
[9] Moura, CP, Cunha, LM, Vilarinho, H et al. Voice parameters in children with Down Syndrome. *J Voice*. 2008; 22(1):34-42. doi:10.1016/j.jvoice.2006.08.011.
[10] Lee, MT, Thorpe, J, Verhoeven, J. Intonation and Phonation in Young Adults with Down Syndrome. *J Voice*. 2009; 23(1):82-87. doi:10.1016/j.jvoice.2007.04.006.
[11] Albertini, G, Bonassi, S, Dall'Armi, V, Giachetti, I, Giaquinto, S, Mignano, M. Spectral analysis of the voice in Down Syndrome. *Res Dev Disabil*. 2010; 31(5):995-1001. doi:10.1016/j.ridd.2010.04.024.

[12] Rodger, R. Voice quality of children and young people with Down's Syndrome and its impact on listener judgement. Queen Margaret University, Edinburgh, Scotland, 2009; Unpublished PhD dissertation. Retrieved from: http://etheses.qmu.ac.uk/352/
[13] Seifpanahi, S, Bakhtiar, M, Salmalian, T. Objective vocal parameters in Farsi-speaking adults with Down syndrome. *Folia Phoniatr et Logop*. 2011; 63(2): 72-76.
[14] Heman-Ackah, YD., Heuer, RJ., Michael, DD, et al. Cepstral peak prominence: A more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003; 112(4):324-333. doi:10.1177/000348940311200406.
[15] Pryce, M. The voice of people with Down's syndrome: An EMG biofeedback study. *Down Syndr Res Pract*. 1994; 2(3):106-111. doi:10.3104/reports.39.
[16] Beck, JM. Organic Variation of the Vocal Apparatus. In: The Handbook of Phonetic Sciences: Second Edition.; 2010: 153-201. doi:10.1002/9781444317251.ch5.
[17] Titze, IR. Workshop on Acoustic Voice Analysis. *Natl Cent Voice Speech, Am*. 1994: 1-36. doi:10.1016/j.jvoice.2009.04.002.
[18] Cavalli, L, Hirson, A. Diplophonia reappraised. *J voice*. 1999; 13(4):542-556. doi:10.1016/S0892-1997(99)80009-5.
[19] Dejonckere, PH, Lebacq, J. An analysis of the diplophonia phenomenon. *Speech Commun*. 1983; 2(1):47-56. doi:10.1016/0167-6393(83)90063-8.
[20] Aronson, AE., Bless, D. Clinical Voice Disorders. 2011; Thieme.
[21] Fawcus, M. The physiology of phonation, in: M. Freeman, M. Fawcus (Eds.), Voice disorders and their management (3rd Ed.), London, Whurr, 1990, pp. 1-17.
[22] Bailly, L, Henrich, N, Pelorson, X. Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling. *J Acoust Soc Am*. 2010; 127(5):3212-3222. doi:10.1121/1.3365220.
[23] Sakakibara, KI., Imagawa, H, Yokonishi, H, Kimura, M, & Tayama, N. Physiological Observations and Synthesis of Subharmonic Voices. *APSIPA ASC*, 2011; 1079–1085.
[24] Stager, S. The Role of the Supraglottic Area in Voice Production. *Otolaryngol.* 2011; S1, 1–7. http://doi.org/10.4172/2161-119X.S1-001
[25] Fourcin, A, Abberton, E. Hearing and phonetic criteria in voice measurement: clinical applications. *Logop Phoniatr Vocology*. 2008; 33(1):35-48. doi:10.1080/14015430701251574.
[26] Schaeffler, F, Beck, JM, Jannetts, S. Phonation stabilisation time as an indicator of voice disorder. 2015; *ICPhS*, Glasgow.
[27] Yin, RK. Designing case studies: Fifth Edition.; 2003; Sage: London
[28] Dunn, LM., & Dunn, DM. *The British picture vocabulary scale*. 1992: GL Assessment Limited.
[29] Shipley, KG., & McAfee, J. Assessment in speech-language pathology. 2009; Delmar Learning: USA.
[30] Teixeira JP., Fernandes PO. Acoustic Analysis of Vocal Dysphonia. In: *Procedia Computer Science*. Vol 64.; 2015:466-473. doi:10.1016/j.procs.2015.08.544.
[31] Boersma, P., Weenink, D. Praat: doing phonetics by computer (Version 5.3.14.)[Computer program]. *http://www.fon.hum.uva.nl/praat/* (accessed 10.01.15).
[32] Yan, Y, Damrose, E, Bless, D. Functional Analysis of Voice Using Simultaneous High-Speed Imaging and Acoustic Recordings. *J Voice*. 2007; 21(5), 604–616. http://doi.org/10.1016/j.jvoice.2006.05.011
[33] Fuks, L. Computer-aided musical analysis of extended vocal techniques for compositional applications. Paper presented at VI Simpósio Brasileiro de Computaçäo Música, Rio de Janeiro. 1999. Available from: http://compmus.ime.usp.br/sbcm/1999/papers/Leonardo_Fuks.pdf. (Accessed 18/12/16)
[34] Kramer, E, Linder, R, Schönweiler, R. A study of subharmonics in connected speech material. *J Voice*. 2013; 27(1):29-38. doi:10.1016/j.jvoice.2012.08.005.
[35] Fourcin, A, McGlashan, J, Blowes, R. Measuring voice in the clinic-Laryngograph® Speech Studio analyses. In Proceedings 6th Voice Symposium of Australia; 2002, October.
[36] Jannetts, S, Schaeffler, F. Cepstral peak prominence-based phonation stabilisation time as an indicator of voice disorder In: PEVOC & MAVEBA 2015, 31st August - 4th September 2015, Florence, Italy. Available from http://eresearch.qmu.ac.uk/3922/ (accessed 02/7/16)
[37] Blomgren, M, Chen, Y, Ng, ML, Gilbert, HR. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *J Acoust Soc Am.* 1998; 103(5), 2649-2658.
[38] Panfili, L. The physiological underpinnings of vowel height and voice quality. *J Acoust Soc Am.* 2016; 139(4), 2221-2221.
[39] Fourcin, A. Aspects of voice irregularity measurement in connected speech. *Folia Phoniatr Logop*. 2009; 61(3):126-136. doi:10.1159/000219948.

[40] Strik, H, Boves, LL B-E. Control of fundamental frequency, intensity and voice quality in speech. *J. Phon*. 1992; 20, 15–25.
[41] Maryn, Y, Corthals, P, Van Cauwenberge, P, Roy, N, De Bodt, M. Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *J Voice*. 2010; 24(5):540-555. doi:10.1016/j.jvoice.2008.12.014.
[42] Baken, RJ, Orlikoff, RF. Clinical measurement of speech and voice. Cengage Learning, New York, 2000.
[43] Fraile, R, Godino-Llorente, JI. Cepstral peak prominence: A comprehensive analysis. *Biomed Signal Process Control*. 2014; 14(1):42-54. doi:10.1016/j.bspc.2014.07.001.
[44] Carding, PN, Steen, IN, Webb, A, Mackenzie, K, Deary, IJ, Wilson, JA. The reliability and sensitivity to change of acoustic measures of voice quality. *Clin Otolaryngol.* 2004; 29(5), 538-544.
[45] Leong, K, Hawkshaw, MJ, Dentchev, D, Gupta, R, Lurie, D, Sataloff, RT. Reliability of objective voice measures of normal speaking voices. *J Voice*. 2013; 27(2):170-176. doi:10.1016/j.jvoice.2012.07.005.
[46] Parsa, V, Jamieson, DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Hearing Res*. 2001; 44(2), 327-339.
[47] Núñez, BF, Suárez, NC, Muñoz, PC, Baragaño, RL, Alvarez, ZM, & Martínez, FA. Spectrographic study of voice disorders: subharmonics. *Acta otorrinolaringologica espanola*. 2000; 51(1), 52-56. Abstract retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10799933
[48] Aichinger P. Diplophonic voice - Definitions, models, and detection. 2015: Graz University of technology, Vienna, Austria; Unpublished PhD thesis. Retrieved from https://www.spsc.tugraz.at/sites/default/files/Thesis_finalVersion.pdf
[49] Aichinger, P, Hagmüller, M, Roesner, I. Bigenzahn, W. Schneider-Stickler, B, & Schoentgen, J. Diplophonia Disturbs Jitter and Shimmer Measurement. *Folia Phoniatrica et Logopaedica*. 2016; 68(1), 22-28.
[50] Bailly, L, Müller, NHBF, Rohlfs, A-K, Hessc, M. Ventricular-Fold Dynamics in Human Phonation. *J Speech, Lang Hear Res*. 2014; 57(June):1679-1691. doi:10.1044/2014.
[51] Sakakibara, K-I, Fuks, L, Imagawa, H, Tayama, N. Growl Voice in Ethic and Pop Styles. In: *Proceedings of the International Symposium on Musical Acoustics (ISMA2004).*; 2004.
[52] Keating, P, Garellek, M, Kreiman, J. Acoustic properties of different kinds of creaky voice. In Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, 2015. Available from: http://www.icphs2015.info/pdfs/Papers/ICPHS0821.pdf. (Accessed 30th May 2016)
[53] Braunschweig, T, Flaschka, J, Schelhorn-Neise, P, Döllinger, M. High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias. Med Eng Phys. 2008; 30(1), 59–66. http://dx.doi.org/10.1016/j.medengphy.2006.12.007
[54] Lee, EK, Son, YI. Muscle tension dysphonia in children: Voice characteristics and outcome of voice therapy. *Int J Pediatr Otorhinolaryngol*. 2005; 69(7):911-917. doi:10.1016/j.ijporl.2005.01.030.
[55] Das, B, Chatterjee, I, Kumar, S. Laryngeal aerodynamics in children with hearing impairment versus age and height matched normal hearing peers. ISRN Otolaryngology, 2013. doi: 10.1155/2013/394604
[56] Coelho, AC, Medved, DM., Brasolotto, AG., in: Prof. Fayez Bahmad (Ed.), Hearing Loss and the Voice, Update On Hearing Loss, InTech, 2015. DOI: 10.5772/61217. Available from: https://www.intechopen.com/books/update-on-hearing-loss/hearing-loss-and-the-voice
[57] Dehqan, A, Scherer, RC. Objective voice analysis of boys with profound hearing loss. *J Voice*, 2015; 25(2), e61-e65.
[58] Alexander, M, Petri, H, Ding, Y, Wandel, C, Khwaja, O, Foskett, N. Morbidity and medication in a large population of individuals with Down syndrome compared to the general population. [Electronic version]. *Dev Med Child Neurology*. 2015; 58(3), 246-254. http://doi.org/10.1111/dmcn.12868
[59] Sun, X. Pitch determination and voice quality analysis using Subharmonic-to-Harmonic Ratio. In: *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol 1.; 2002:I - 333 - I - 336. doi:10.1109/ICASSP.2002.5743722.

M1: Demographic data

Age (years; months)	23;9	**Perceived voice quality**	loud, gruff and harsh; low-pitched
Verbal mental age (BPVS)	3;4 (2;11 – 3;9)	**Pitch range of vocal glides, (min-max f0), Hz**	108.94 -238.84
Hearing impairment	Unconfirmed	**Mean f0 and range (comment to peer), Hz**	97.05 (50.29 -193.71)
		Mean f0 and range (picture description), Hz	118.57 (53.71-221.21)

M1 Voice data

Vowel	**target pitch, Hz**	**Mean f_0, Hz**	**HNR, dB**	**Jitter, %**	**Shimmer, %**	**Duration of vowel (T), ms**	**SH, ms at onset (t= 100 ms)**	**SH, ms at medially**	**SH, ms at offset (t= 100 ms)**	**Percentage duration of SH:whole vowel duration, %**	**Diplophonia**
/a/1	131	97.80	5.96	1.58	9.53	876	100*	+676*+	100*	100	Yes
/a/2	165	155.66	**22.29**	**0.31**	**2.88**	1405	49	16+	100	1.14	No
/a/3	196	177.50	**23.80**	**0.23**	**2.61**	1896	100	+131 32+	100	9.61	No
Mean /a/			17.35	0.71	5.01						
/i/1	131	212.11	**23.73**	0.89	**2.89**	1040	39	0	90	12.40	No
/i/2	165	180.68	**20.86**	0.89	**2.04**	1450	93	49+	100	16.69	No
/i/3	165	212.03	**23.77**	0.88	**2.90**	950	97	10+	100	21.79	No
Mean /i/			**22.78**	0.88	**2.61**						
/u/1	196	229.78	**28.35**	**0.47**	**2.11**	688	72	7+	100	26.02	No
/u/2	131	162.99	**26.01**	0.91	**2.65**	870	100	+16	91	26.53	No
/u/3	131	175.31	**24.12**	0.84	**2.15**	915	93	0	91	20.11	No
Mean /u/			**26.16**	0.74	**2.30**						

Table 1. Demographic and voice data for Male 1 (M1).

The top section of the table shows a summary of M1's performance on key cognitive tasks, and existing information on his hearing abilities and voice characteristics. In both speaking tasks, his minimum pitch was within vocal fry range. The British Picture Vocabulary Scale 1(BPVS) was used to assess verbal mental age: the table shows the mental age equivalent and the age range (years; months). The second section provides data derived from numerical acoustic analysis of medial portions (T-[onset +offset]) of his sustained vowels (mean f0, Hz; HNR, dB; Jitter, %; Shimmer, %). Values in bold type indicate results that are within norms for HNR, jitter and shimmer, based on ranges reported by Teixeira and Fernandes (2015):

/u/: (HNR: 22.319-31.261 dB; jitter: 0.183-0.517 %; shimmer: 0.13-6.23 %)

/i/: (HNR: 20.52-27.32 dB; jitter: 0.196-0.444 %; shimmer: 1.254-3.106 %)

/a/: (HNR: 21.50-25.80 dB; jitter: 0.234-0.446 %; shimmer 1.811-3.469 %)

The presence and duration of visible subharmonics (SH) at 35 dB is expressed in ms, and as a percentage of the whole vowel (T). Subharmonic data are provided for each segment of the sustained vowel (onset, medially, offset). The + indicates where medial subharmonics are continuous with onset (+t) or with offset (t+); * marks sections of vowel that were perceived as diplophonic. The incidence of perceived diplophonia is indicated (Yes or No).

Means are given for HNR jitter and shimmer, and for average duration of subharmonics at onset, medially, and offset; and average percentage duration of subharmonics, as a proportion of the whole vowel.

M2: Demographic data

Age (years;months)	23;10	**Perceived voice quality**	Breathy and pubophonic
Verbal mental age (BPVS)	3;0 (2;8 - 3;5)	**Pitch range of vocal glides, (min-max f0), Hz**	163.09 -421.82
Hearing impairment	Moderate bilateral loss	**Mean f0 and range (comment to peer), Hz**	178.09 (154.59 220.24)
		Mean f0 and range (picture description), Hz	172.93 (158.09-197.37)

M2: Voice data

Vowel	**Target pitch, Hz**	**Mean f_0, Hz**	**HNR, dB**	**Jitter, %**	**Shimmer, %**	**Duration of vowel (T), ms, and Overlap at onset, ms**	**SH, ms at onset (t= 100 ms)**	**SH, ms at medially**	**SH, ms at offset (t= 100 ms)**	**Percentage duration of SH: whole vowel duration (T-Overlap), %**	**Diplophonia**
/a/1	262	174.24	19.35	0.64	3.84	103/2190	n/k	116*, 126*, 77, 107	0	n/k	Yes
/a/2	330	261.54	21.03	0.45	4.17	125/1542	n/k	168*	0	n/k	Yes
/a/3	196	287.56	**23.08**	**0.28**	3.57	331/1228	n/k	0	0	n/k	No
Mean /a/			21.15	0.45	3.86						
/i/1	262	245.21	**25.03**	0.46	**3.11**	236/2250	n/k	59, 73	0	6.55	No
/i/2	262	301.23	**23.94**	0.47	3.82	1086	100	+2	0	9.4	No
/i/3	330	276.94	**25.43**	0.54	**2.99**	152/1135	n/k	0	0	n/k	No
Mean /i/			**24.80**	0.49	3.31						
/u/1	262	238.98	**25.96**	**0.39**	**3.75**	604/1088	n/k	n/k	0	n/k	No
/u/2	330	263.91	21.66	**0.44**	**3.89**	1272	100	+252	0	27.67	No
/u/3	196	271.32	**24.97**	**0.39**	**2.22**	1005	100	+126	0	22.48	No
Mean/u/			**24.19**	**0.41**	**3.28**						

Table 2. Demographic and voice data for Male 2 (M2).

The top section of the table shows a summary of M2's performance on key cognitive tasks, and existing information on his hearing abilities and voice characteristics. The British Picture Vocabulary Scale 1(BPVS) was used to assess verbal mental age: the table shows the mental age equivalent and the age range (years; months). The second section provides data derived from numerical acoustic analysis of medial portions (T- [onset +offset]) of his sustained vowels (mean f0, Hz; HNR, dB; Jitter, %; Shimmer, %). The onset of six vowels overlapped with the demonstration; the duration of overlap at onset is shown with the duration of vowel (ms). The overlap made it impossible to confirm the incidence or duration of subharmonics or diplophonia at onset: values that are not known (n/k) are indicated. Values in bold type indicate results that are within norms for HNR, jitter and shimmer, based on ranges reported by Teixeira and Fernandes (2015):

/u/: (HNR: 22.319-31.261 dB; jitter: 0.183-0.517 %; shimmer: 0.13-6.23 %)

/i/: (HNR: 20.52-27.32 dB; jitter: 0.196-0.444 %; shimmer: 1.254-3.106 %)

/a/: (HNR: 21.50-25.80 dB; jitter: 0.234-0.446 %; shimmer 1.811-3.469 %)

The presence and duration of visible subharmonics (SH) at 35 dB is expressed in ms, and as a percentage of the whole vowel (T). Subharmonic data are provided for each segment of the sustained vowel (onset, medially, offset). The + indicates where medial subharmonics are continuous with onset (+t) or with offset (t+); * marks sections of vowel that were perceived as diplophonic. The incidence of perceived diplophonia is indicated (Yes or No).

Means are given for HNR jitter and shimmer, and for average duration of subharmonics at onset, medially, and offset; and average percentage duration of subharmonics, as a proportion of the whole vowel.

F1: Demographic data

Age (years; months)	23;4	**Perceived voice quality**	Harsh, gruff and breathy
Verbal mental age (BPVS)	3;1 (2;9 - 3;7 years)	**Pitch range of vocal glides, (min-max f0), Hz**	221.20 - 324.10
Hearing impairment	Unconfirmed	**Mean f0 and range (comment to peer), Hz**	281.42 (213.96 -358.60)
		Mean f0 and range (picture description), Hz	245.55 (206.08-293.27)

F1: Voice data

Vowel	**Target pitch, Hz**	**Mean f_0, Hz**	**HNR, dB**	**Jitter, %**	**Shimmer, %**	**Duration of vowel (T), ms, and Overlap at onset, ms,**	**SH, ms at onset (t= 100 ms)**	**SH, ms at medially**	**SH, ms at offset (t= 100 ms)**	**Percentage duration of SH:whole vowel duration, %**	**Diplophonia**
/a/1	330	205.56	16.65	0.62	6.30	660	100	89*, 225*	0	62.72	Yes
/a/2	261	248.11	21.33	0.52	**3.47**	529	100	+29	0	24.38	No
/a/3	261	238.40	**22.75**	**0.58**	**3.11**	99/420	n/k	230*	100*	100	Yes
/a/4	261	227.14	**21.72**	0.55	**3.02**	707	67	0	0	9.48	No
Mean /a/			20.61	0.57	3.97						
/i/1	330	298.53	20.12	0.76	4.61	558	0	37	27	11.47	No
/i/2	292	270.82	**22.84**	0.68	3.41	574	100	+18, 76	0	33.80	No
/i/3	292	142.63	12.17	1.78	4.43	510	100	+122, 88*,	100	80.39	Yes
/i/4	330	293.19	**21.60**	**0.27**	4.09	282/842	n/k	284*, 98+	68	80.35	Yes
Mean /i/			19.18	0.87	4.13						
/u/1	330	249.48	**27.59**	**0.47**	**2.06**	553	100	71	0	30.92%	No
/u/2	261	174.39	**24.75**	**0.31**	**2.94**	561	100	+18, 332*	0	80.21%	Yes
/u/3	261	286.07	**27.49**	**0.26**	**2.07**	294/471	n/k	n/k	0	n/k	No
Mean /u/			**26.61**	**0.34**	**2.36**						

Table 3. Demographic and voice data for Female 1 (F1).

The top section of the table shows a summary of F1's performance on key cognitive tasks, and existing information on her hearing abilities and voice characteristics. The British Picture Vocabulary Scale 1(BPVS) was used to assess verbal mental age: the table shows the mental age equivalent and the age range (years; months). The second section provides data derived from numerical acoustic analysis of medial portions (T- [onset +offset]) of her sustained vowels (mean f0, Hz; HNR, dB; Jitter, %; Shimmer, %). The onset of three vowels overlapped with the demonstration; the duration of overlap at onset is shown with the duration of vowel (ms). The overlap made it impossible to confirm the incidence or duration of subharmonics or diplophonia at onset: values that are not known (n/k) are indicated. Values in bold type indicate results that are within norms for HNR, jitter and shimmer, based on ranges reported by Teixeira and Fernandes (2015):

/u/: (HNR: 22.319-31.261 dB; jitter: 0.183-0.517 %; shimmer: 0.13-6.23 %)

/i/: (HNR: 20.52-27.32 dB; jitter: 0.196-0.444 %; shimmer: 1.254-3.106 %)

/a:/ (HNR: 21.50-25.80 dB; jitter: 0.234-0.446 %; shimmer 1.811-3.469 %)

The presence and duration of visible subharmonics (SH) at 35 dB is expressed in ms, and as a percentage of the whole vowel (T). Subharmonic data are provided for each segment of the sustained vowel (onset, medially, offset). The + indicates where medial subharmonics are continuous with onset (+t) or with offset (t+); * marks sections of vowel that were perceived as diplophonic. The incidence of perceived diplophonia is indicated (Yes or No).

Means are given for HNR jitter and shimmer, and for average duration of subharmonics at onset, medially, and offset; and average percentage duration of subharmonics, as a proportion of the whole vowel.

F2: Demographic Data			
Age (years;months)	23;10	**Perceived voice quality**	Some roughness, slightly breathy
Verbal mental age (BPVS)	3;0 (2;8 - 3;5)	**Pitch range of vocal glides, (min-max f0), Hz**	152.40 - 472.40
Hearing impairment	Unconfirmed	**Mean f0 and range (comment to peer), Hz**	245.28 (130.45- 356.57)
		Mean f0 and range (picture description), Hz	215.12 (110.31-295.32)

F2: Voice Data

Vowel	**Target pitch, Hz**	**Mean f_0, Hz**	**HNR, dB**	**Jitter, %**	**Shimmer, %**	**Duration of vowel (T), ms, and Overlap at onset, ms**	**SH, ms at onset (t= 100 ms)**	**SH, ms at medially**	**SH, ms at offset (t= 100 ms)**	**Percentage duration of SH:whole vowel duration, %**	**Diplophonia**
/a/1	261	281.51	19.19	0.51	4.15	1041	100	104	0	19.59	No
/a/2	261	257.92	**22.07**	**0.34**	**2.33**	1105	100	53	0	13.85	No
/a/3	330	295.08	**22.06**	**0.30**	**3.32**	1453	100	56	0	10.74	No
/a/4	330	306.03	15.71	0.59	14.33	975	100	+69, 622*	0	81.13	Yes
Mean /a/			19.75	**0.44**	6.03		100	180.80	0	31.32	1
/i/1	330	332.02	**26.65**	**0.23**	**1.38**	1387	100	+7	0	7.71	No
/i/2	261	267.23	**26.76**	**0.29**	**1.67**	974	100	+6	0	10.88	No
Mean /i/			**26.70**	**0.26**	**1.52**		100	6.5	0	9.29	0
/u/1	330	300.64	**27.91**	**0.38**	**2.24**	1264	100	+49, 182	0	26.19	No
/u/2	261	269.51	**29.26**	**0.22**	**1.85**	1849	96	0	0	5.19	No
/u/3	261	253.29	**31.18**	**0.15**	**1.75**	691/795	n/k	n/k	0	n/k	No
Mean /u/			**29.45**	**0.25**	**1.95**		65.33	77.00	0	16.04	0

Table 4. Demographic and voice data for Female 2 (F2). The top section of the table shows a summary of his performance on key The top section of the table shows a summary of F2's performance on key cognitive tasks, and existing information on her hearing abilities and voice characteristics. The British Picture Vocabulary Scale 1(BPVS) was used to assess verbal mental age: the table shows the mental age equivalent and the age range (years; months). The second section provides data derived from numerical acoustic analysis of medial portions (T- [onset +offset]) of her sustained vowels (mean f0, Hz; HNR, dB; Jitter, %; Shimmer, %). The onset of one vowel overlapped with the demonstration; the duration of overlap at onset is shown with the duration of vowel (ms). The overlap made it impossible to confirm the incidence or duration of subharmonics or diplophonia at onset: values that are not known (n/k) are indicated. Values in bold type indicate results that are within norms for HNR, jitter and shimmer, based on ranges reported by Teixeira and Fernandes (2015):

/u/: (HNR: 22.319-31.261 dB; jitter: 0.183-0.517 %; shimmer: 0.13-6.23 %)

/i/: (HNR: 20.52-27.32 dB; jitter: 0.196-0.444 %; shimmer: 1.254-3.106 %)

/a/: (HNR: 21.50-25.80 dB; jitter: 0.234-0.446 %; shimmer 1.811-3.469 %)

The presence and duration of visible subharmonics (SH) at 35 dB is expressed in ms, and as a percentage of the whole vowel (T). Subharmonic data are provided for each segment of the sustained vowel (onset, medially, offset). The + indicates where medial subharmonics are continuous with onset (+t) or with offset (t+); * marks sections of vowel that were perceived as diplophonic. The incidence of perceived diplophonia is indicated (Yes or No).

Means are given for HNR jitter and shimmer, and for average duration of subharmonics at onset, medially, and offset; and average percentage duration of subharmonics, as a proportion of the whole vowel.

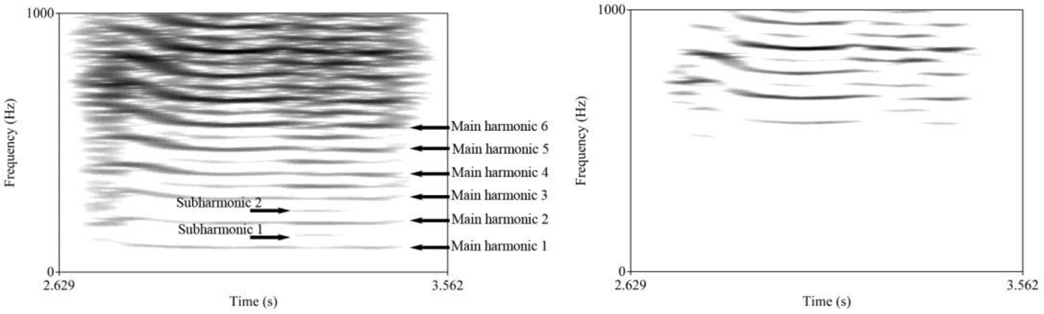

Figure 1: Narrow-band spectrogram of M1s sustained /a/ vowel at 35 dB (a: left) and 15 dB (b: right). The arrows on the image at 35 dB indicate the presence and position of the first two single subharmonics and the first six main harmonics. At 15 dB, the upper subharmonics and main harmonics remain visible in the 500-1000 Hz range only.

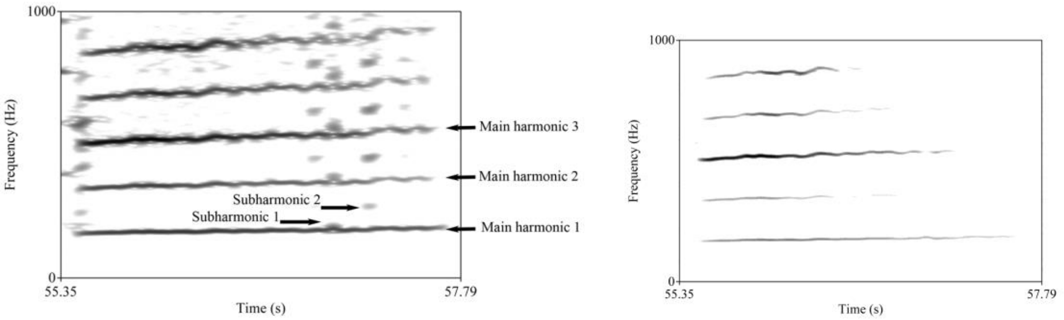

Figure 2: Narrow-band spectrogram of M2's sustained /a/ vowel at 35 dB (a: left) and 15 dB (b: right). The arrows on the image at 35 dB indicate the presence and position of the first two single subharmonics and first three main harmonics. The subharmonics at onset interact with the offset of the demonstrated vowel. At 15 dB, no subharmonics remain visible.

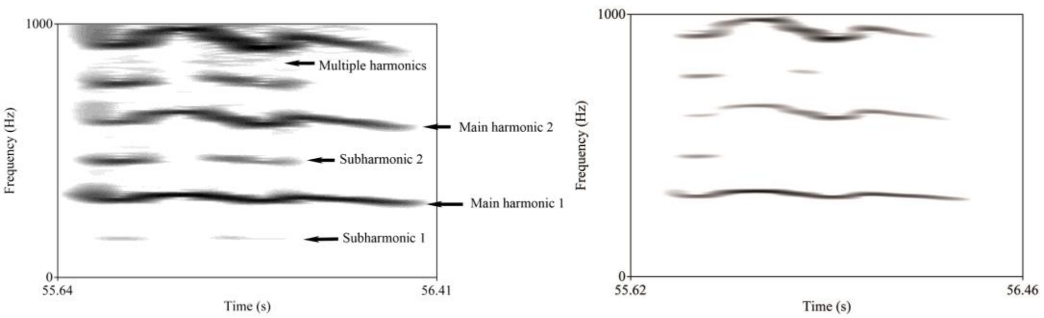

Figure 3: Narrow-band spectrogram of F1's sustained /a/ vowel at 35 dB (a: left) and 15 dB (b: right). The arrows on the image at 35 dB indicate the presence and position of the first two single subharmonics and main harmonics, and show the presence of multiple harmonics in the upper frequency range. At 15 dB, the subharmonics are visible at onset in the mid frequency range, and in the upper frequency range during in the medial portion of the vowel.

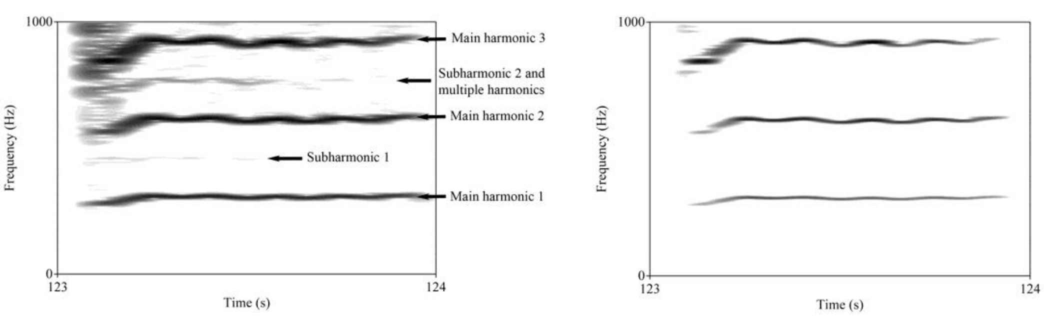

Figure 4: Narrow-band spectrogram of F2's sustained /a/ vowel at 35 dB (a: left) and 15 dB (b: right). The arrows on the image at 35 dB indicate the presence and position of the first two single subharmonics and three main harmonics. There

are and multiple harmonics in the upper frequency range, above subharmonic 2. At 15 dB, only the subharmonics above main harmonic 3 remain visible at onset.