ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Speech Communication xxx (2011) xxx–xxx

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility

Tiago H. Falk [a,*], Wai-Yip Chan [b], Fraser Shein [c,d]

[a] *Institut National de la Recherche Scientifique, INRS-EMT, Montréal, Canada*
[b] *Department of Electrical and Computer Engineering, Queen's University, Kingston, Canada*
[c] *Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada*
[d] *Department of Computer Science, University of Toronto, Toronto, Canada*

## Abstract

Objective measurement of dysarthric speech intelligibility can assist clinicians in the diagnosis of speech disorder severity as well as in the evaluation of dysarthria treatments. In this paper, several objective measures are proposed and tested as correlates of subjective intelligibility. More specifically, the kurtosis of the linear prediction residual is proposed as a measure of vocal source excitation oddity. Additionally, temporal perturbations resultant from imprecise articulation and atypical speech rates are characterized by short- and long-term temporal dynamics measures, which in turn, are based on log-energy dynamics and on an auditory-inspired modulation spectral signal representation, respectively. Motivated by recent insights in the communication disorders literature, a composite measure is developed based on linearly combining a salient subset of the proposed measures with conventional prosodic parameters. Experiments with the publicly-available 'Universal Access' database of spastic dysarthric speech (10 patient speakers; 300 words spoken in isolation, per speaker) show that the proposed composite measure can achieve correlation with subjective intelligibility ratings as high as 0.97; thus the measure can serve as an accurate indicator of dysarthric speech intelligibility.

*Keywords:* Dysarthria; Vocal source excitation; Temporal dynamics; Intelligibility; Linear prediction

## 1. Introduction

Dysarthria comprises a group of motor speech disorders resultant from damage to the central and/or peripheral nervous systems (Doyle et al., 1997). Dysarthric speech is often associated with excessive nasalization, disordered speech prosody, imprecise articulation, and variable speech rate (Doyle et al., 1997) – factors that often render speech unintelligible. One of the most common subtypes of dysarthria is termed "spastic dysarthria" with symptoms that can range from strained phonation, imprecise placement of articulators, incomplete consonant closure, monotone speech, and reduced voice onset time distinctions between voiced and unvoiced stops (Duffy, 2005). Spastic dysarthria is most commonly associated with cerebral palsy and traumatic brain injury (Duffy, 2005).

Currently, speech-language pathologists mainly rely on subjective intelligibility assessment tests to characterize the severity of speech disorders, as well as to monitor, plan treatment, and document changes in intelligibility over time (Klopfenstein, 2009). Subjective intelligibility tests, however, are costly, laborious, and subject to many intrinsic variables and biases due to e.g., familiarity with the patients and their speech pathologies (De Bodt et al., 2002; Van Nuffelen et al., 2009). Objective measurement, on the other hand, is economical and reliable (repeatable) and can assist in surgical and/or pharmacological treatment evaluation as well as in remote patient rehabilitation monitoring

[*] Corresponding author. Tel.: +1 514 875 1266x3066; fax: +1 514 875 0344.
*E-mail address:* tiago.falk@ieee.org (T.H. Falk).

0167-6393/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2011.03.007

(Constantinescu et al., 2010). In fact, there is growing evidence suggesting that clinicians are becoming more receptive to automated machine-based systems that assist in treatment decisions (e.g., Hill et al., 2006; Maier et al., 2009).

In the past, a handful of objective intelligibility measures have been proposed for dysarthric speech. The system proposed by Middag et al. (2009) used phonemic and phonological features that were force-aligned to the acoustic-phonetic transcription of the target word. Alignment was achieved by means of an automatic speech alignment algorithm trained on acoustic models of "healthy" speech. Features were then mapped to an intelligibility score using a linear regression function. Additionally, the work described by Gu et al. (2005) computed distance measures (e.g., Itakura–Saito distortion) between the produced disordered speech utterance and the same utterance spoken by a healthy individual. To account for differences in utterance durations, dynamic time warping was applied.

Today, automatic speech recognition (ASR) has become a popular method of objectively quantifying dysarthric speech intelligibility for speakers with mild or moderate dysarthria (e.g., Doyle et al., 1997; Ferrier et al., 1995; Maier et al., 2009; Sharma et al., 2009); technological advances, however, are still needed before ASR is used for severe dysarthric speakers (Middag et al., 2009; Rudzicz, 2007). Major limiting factors in the widespread use of ASR, however, include limited vocabulary sizes ranging from 10–70 words (Doyle et al., 1997), the need for speaker-dependent (or adaptive) acoustic models (Raghavendra et al., 2001; Rudzicz, 2007), and the sparseness of available data needed to accurately train such models (Green et al., 2003).

The methods mentioned above require *a priori* information, such as the signal or feature prototypes of the target word being uttered. In many practical applications, however, such information may not be available and "blind" measures are more convenient. The majority of existing blind methods rely on prosodic measurements, such as fundamental frequency ($f0$) variation, tone unit duration, and second-formant slope transitions (Bunton et al., 2000; Schlenck et al., 1993; Kent et al., 1989), which have been shown to be useful indicators of dysarthric speech intelligibility (Klopfenstein, 2009). Recently, the power spectrum of the envelope of the speech signal, or modulation spectrum, was used to characterize rhythmic disturbances in dysarthric speech. The study suggested that the perturbations of speech temporal patterns associated with dysarthria played an important role in intelligibility (LeGendre et al., 2009).

Subjective listening tests of dysarthric speech suggest that intelligibility can be expressed as a weighted linear combination of different perceptual dimensions, such as articulation, vocal harshness, prosody, and nasality (De Bodt et al., 2002). In this paper, several parameters are proposed and tested as correlates of subjective intelligibility. The parameters measure abnormal behaviours found in dysarthric speech, such as vocal source excitation oddity, temporal dynamics perturbations, hypernasality, and disordered prosody. Our results (see Section 3.3) suggest that the measures are complementary and when linearly combined can serve as an accurate indicator of dysarthric speech intelligibility. Moreover, in comparison with ASR, the proposed method is considerably simpler to design and implement. The remainder of this paper is organized as follows: Section 2 describes the proposed measures; experimental setup and results are reported in Section 3; and discussion and conclusions are presented in Sections 4 and 5, respectively.

## 2. Objective measurement of spastic dysarthric word intelligibility

Several factors are known to adversely affect speech intelligibility for individuals with dysarthria (De Bodt et al., 2002). The most prominent are associated with atypical vocal source excitation (e.g., vocal harshness), temporal dynamics (e.g., unclear distinction between adjacent phonemes due to imprecise placement of articulators), hypernasality, and disordered prosody (e.g., monotonicity). In order to blindly assess speech intelligibility, measures need to be developed such that perturbations in typical vocal source excitation, temporal dynamics, nasality, and prosody can be characterized. In this paper, several such measures are proposed and tested as correlates of subjective intelligibility.

### 2.1. Separation of vocal source and vocal tract information

Linear prediction (LP) analysis has been widely used in speech applications to separate vocal source (glottal) excitation, $u(n)$, and vocal tract modulation, $h(n)$, from the produced speech signal, $s(n) = u(n) * h(n)$, where "*" indicates convolution (Benesty et al., 2008). Commonly, the vocal tract is modeled as a time-varying all-pole filter given by

$$H(z) = \frac{1}{A(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}. \qquad (1)$$

Hence, the produced speech signal $s(n)$ can be approximated by

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n). \qquad (2)$$

The coefficients $a_k$, $k = 1, \ldots, p$, of the all-pole filter depend on the shape and resonant characteristics of the vocal tract and determine the spectral characteristics of the particular sound being generated. The excitation signal, in turn, is approximated either as a quasi-periodic train of impulses for voiced speech segments, random noise for unvoiced segments, or a combination thereof for voiced fricatives (e.g., 'v') (Benesty et al., 2008); the multiplicative factor $G$ is the gain applied to the excitation signal.

Linear prediction analysis assumes that the current signal sample can be predicted by a linear combination of $p$ previous samples. The predicted sample $\hat{s}(n)$ is given by

Please cite this article in press as: Falk, T.H. et al., Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, Speech Comm. (2011), doi:10.1016/j.specom.2011.03.007

$$\hat{s}(n) = \sum_{k=1}^{p} \hat{a}_k s(n-k). \quad (3)$$

With this assumption, the linear prediction error, or LP residual $r(n)$, will correspond to the excitation signal multiplied by the gain factor $G$,

$$r(n) = s(n) - \hat{s}(n) \cong Gu(n), \quad (4)$$

where coefficients $\hat{a}_k$ in (3) can be estimated by minimizing the energy of $r(n)$ via an autocorrelation or a covariance method (Benesty et al., 2008). Commonly, $p = 10$ and $p = 18$ are used for speech sampled at 8 and 16 kHz sample rates, respectively. In summary, with LP analysis, vocal source information is characterized by the LP residual signal and vocal tract shaping is characterized by the estimated LP parameters $\hat{a}_k$. In the results described herein, LP analysis is performed over 32-millisecond frames with 10-millisecond frame shifts.

### 2.2. Atypical vocal source excitation

As seen from (4), linear prediction residuals correspond to the vocal source excitation signals, thus for voiced speech segments, are associated with glottal pulses and have strong impulse-like peaks (Ananthapadmanabha and Yegnanarayana, 1979). For clean natural speech, such LP residual peaks will render the LP residual distribution with a heavier tail or a higher kurtosis value (Gillespie et al., 2001). With spastic dysarthric speech, however, its vocal harshness characteristics are produced by more prominent noise-like excitation patterns, which are associated with a decrease in LP residual kurtosis values. This behaviour is illustrated by the plots in Fig. 1, where subplots (a) and (b) show, from top to bottom, the waveform and LP residual signals for severe and mild dysarthric speech, respectively.

As can be seen from the plots, for the severe dysarthric speech signal (with subjective word intelligibility around 6%), the LP residual signal exhibits prominent noise-like characteristics with a kurtosis value approaching nullity (i.e., that of a Gaussian distribution). Additionally, the monotone sinusoidal-like characteristics of the speech signal can be more accurately represented by a linear prediction model, causing a decrease in the residual signal peaks. As intelligibility increases, however, a more typical LP residual pattern can be observed with prominent peaks occurring during glottal excitations (see Fig. 1(b)), causing an increase in the LP residual kurtosis. In order to quantify such vocal source excitation atypicality, an LP residual kurtosis metric $\mathcal{K}$ is used

$$\mathcal{K}_{LP} = \frac{N\sum_{n=1}^{N}(r(n)-\bar{r})^4}{\left(\sum_{n=1}^{N}(r(n)-\bar{r})^2\right)^2} - 3, \quad (5)$$

where $\bar{r}$ indicates the sample average of $r(n)$ and $N$ is the total number of sample points.

### 2.3. Perturbations in temporal dynamics

Speech temporal impairments can include unclear distinction between adjacent phonemes due to imprecise placement of articulators, slower speech rates, and rhythmic disturbances, to name a few (Duffy, 2005). LeGendre et al. (2009) used long-term temporal dynamics (256 ms and greater) to characterize rhythm pattern perturbations. In this paper, both short-term and long-term temporal dynamics measures are computed and explored as possible indicators of dysarthric word intelligibility.

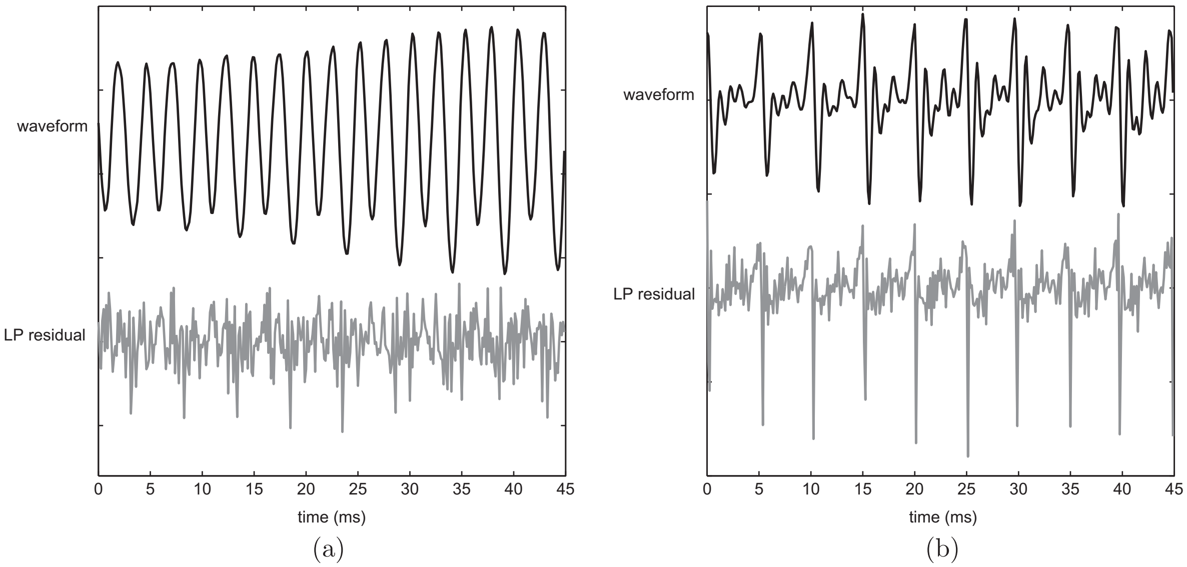

Fig. 1. Waveform and LP residual of a 45 ms segment of the phoneme /o/ in ‘for’. The subfigures are for individuals with (a) severe dysarthria and very low word intelligibility (6%) and (b) mild dysarthria and very good intelligibility (95%).

### 2.3.1. Short-term temporal dynamics

Here, the zeroth order cepstral coefficient ($c_0 = \log G$) is computed as a measure of short-term log-spectral energy and the zeroth order delta coefficient is used as a measure of log-energy rate of change (Huang et al., 2001). Let $c_0(m)$ denote the zeroth order cepstral coefficient for frame $m$. $\Delta c_0(m)$ represents the zeroth order delta coefficient and is computed as Picone (1993)

$$\Delta c_0(m) = \sum_{l=-L}^{L} l\ c_0(m+l), \tag{6}$$

where the normalization factor $\sum_{l=-L}^{L} l^2$ is omitted as it does not affect the results and $L = 3$ is used.

The plots in Fig. 2(a) and (b) depict, from top to bottom, the waveform, log-energy, and delta log-energy for severe and mild dysarthric speech utterances of the word ‘overshadowed’, respectively. In order to capture temporal dynamics information, sample statistics are computed from $C$ samples of $\Delta c_0$ represented by $x_i$ below. In particular, the standard deviation ($\sigma_\Delta$), skewness ($\mathcal{S}_\Delta$), and kurtosis ($\mathcal{K}_\Delta$) are computed according to

$$\sigma_\Delta = \sqrt{\frac{1}{C-1}\sum_{i=1}^{C}(x_i - \bar{x})^2}, \tag{7}$$

$$\mathcal{S}_\Delta = \frac{\sqrt{C}\sum_{i=1}^{C}(x_i - \bar{x})^3}{\left(\sum_{i=1}^{C}(x_i - \bar{x})^2\right)^{3/2}}, \tag{8}$$

$$\mathcal{K}_\Delta = \frac{C\sum_{i=1}^{C}(x_i - \bar{x})^4}{\left(\sum_{i=1}^{C}(x_i - \bar{x})^2\right)^2} - 3, \tag{9}$$

where $\bar{x}$ indicates the sample average of $x_i$. The skewness and kurtosis parameters are used as a measure of asymmetry and peakedness of the distribution of the $\Delta c_0$ samples, respectively.

### 2.3.2. Long-term temporal dynamics

In order to capture long-term temporal dynamics of the speech signal, an auditory-inspired modulation spectral signal representation is used (Falk and Chan, 2010). The modulation spectrum characterizes the rate of change of long-term speech temporal envelopes. In our experiments, the modulation spectral signal representation is obtained using the signal processing steps depicted by Fig. 3.

First, the dysarthric speech signal $s(n)$ is filtered by a bank of 23 equivalent rectangular bandwidth critical-band filters (Glasberg and Moore, 1990; Slaney, 1993). The output signal of the $j$th filter is given by

$$s_j(n) = s(n) * h_j(n), \tag{10}$$

where $h_j(n)$ is the $j$th filter impulse response. Temporal dynamics information is obtained from the temporal envelope of $s_j(n)$. The temporal envelope (or Hilbert envelope) is given by the magnitude of the complex analytic signal, namely, $\tilde{s}_j(n) = s_j(n) + \jmath\mathcal{H}\{s_j(n)\}$, where $\mathcal{H}\{\cdot\}$ denotes the Hilbert transform. Hence,

$$e_j(n) = \sqrt{s_j(n)^2 + \mathcal{H}\{s_j(n)\}^2}. \tag{11}$$

Temporal envelopes $e_j(n)$ are then windowed, with the $m$th windowed $e_j(n)$ denoted henceforth as $e_j(m)$, where the time variable $n$ is dropped for convenience. The discrete Fourier transform $\mathcal{F}\{\cdot\}$ is then used to compute the modulation spectrum $E_j(m;f) = |\mathcal{F}(e_j(m))|$ for frame $m$ and modulation frequency $f$. Lastly, modulation frequency bins are grouped into $K$ bands. In order to emulate an auditory-inspired modulation filterbank (Dau et al., 1996), $K = 8$ second-order bandpass filters with a quality factor $Q = 2$

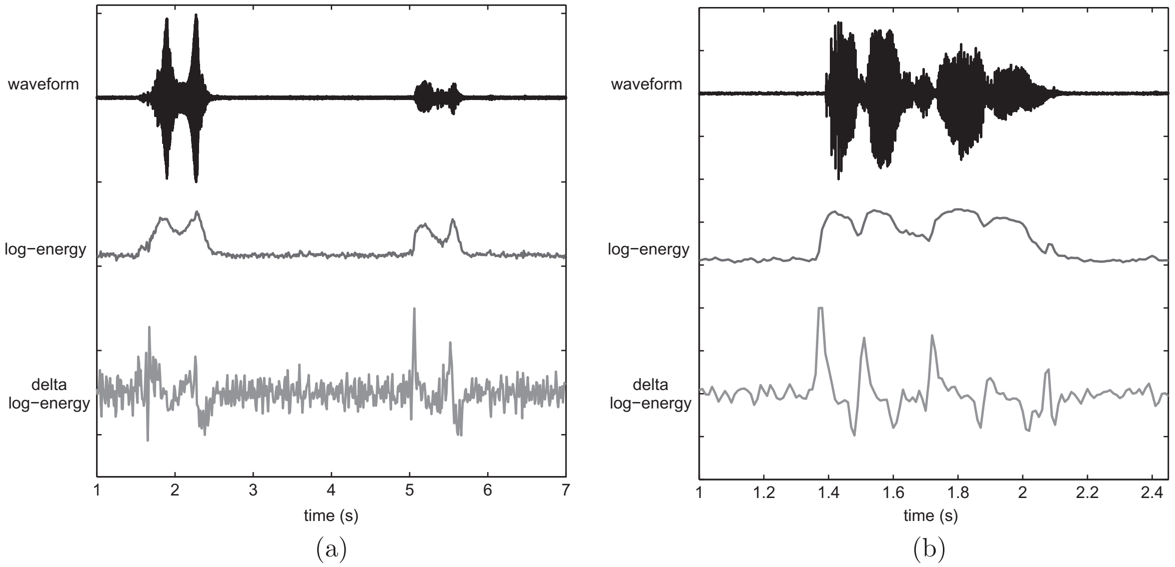

Fig. 2. Waveform, log-energy, and delta log-energy for individuals with (a) severe dysarthria and very low word intelligibility (6%) and (b) mild dysarthria and very good intelligibility (95%). The word being uttered is ‘overshadowed’.

Please cite this article in press as: Falk, T.H. et al., Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, Speech Comm. (2011), doi:10.1016/j.specom.2011.03.007

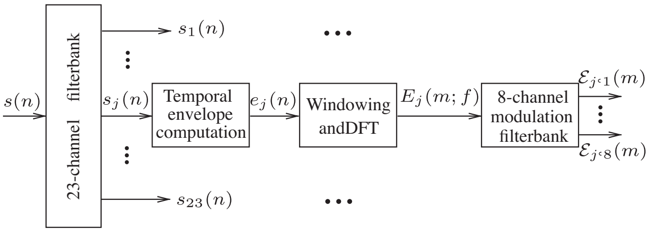

Fig. 3. Block diagram of the signal processing steps involved in the computation of the modulation spectrum.

are used; filter center frequencies range from 2–64 Hz. The $k$th modulation band energy for frame $m$ is denoted as $\mathcal{E}_{j,k}(m),\ k = 1,\ldots,K$.

The modulation energy $\mathcal{E}_{j,k}(m)$ is then averaged over all active speech frames to obtain

$$\bar{\mathcal{E}}_{j,k} = \frac{1}{N_{act}} \sum_{i=1}^{N_{act}} \mathcal{E}^{act}_{j,k}(i), \qquad (12)$$

where $N_{act}$ denotes the number of active speech frames (found via a simple energy-threshold based voice activity detection algorithm) and $\mathcal{E}^{act}_{j,k}(i)$ the modulation energy of such frames. Previous research has suggested that natural speech contains dominant modulation frequencies from 2–20 Hz (Drullman et al., 1994b,a) with spectral peaks at approximately 4 Hz (Arai et al., 1996). It is hypothesized that prolonged phonemes, slower speech rates, as well as the unclear distinction between adjacent phonemes caused by imprecise placement of articulators, will cause a shift of the modulation frequency content to modulation frequencies below 4 Hz. In turn, as intelligibility levels increase, modulation frequency content will be better spread across higher modulation frequencies, as observed with natural speech (Drullman et al., 1994a). In order to characterize this oddity in speech temporal dynamics, the ratio of modulation spectral energy at modulation frequencies lesser than 4 Hz to modulation frequencies greater than 4 Hz is proposed. The low-to-high modulation energy ratio (LHMR) is given by

$$\text{LHMR} = \frac{\sum_{k=1}^{\mathfrak{R}}\sum_{j=1}^{23}\overline{\mathcal{E}}_{j,k}}{\sum_{k=\mathfrak{R}+1}^{8}\sum_{j=1}^{23}\overline{\mathcal{E}}_{j,k}}, \qquad (13)$$

where $\mathfrak{R}$ corresponds to the index of the modulation filter centered at approximately 4 Hz; in our simulations, this corresponds to $\mathfrak{R} = 4$.

The proposed LHMR measure differs from the one proposed by LeGendre et al. (2009) (called Ratio_2000) in several manners. First, the developed measure incorporates information across all 23 acoustic frequency bands and not just the octave band centered at 2 kHz. Second, the use of the Hilbert transform for temporal envelope calculation, as opposed to half-wave rectification and lowpass filtering at 30 Hz, allows for modulation frequencies beyond 30 Hz to be incorporated; such higher frequencies have been shown in the past to be important for intelligibility estimation (Falk et al., 2010; Drullman et al., 1994a). Lastly, the proposed measure emulates the psychoacoustic insights described by Dau et al. (1996, 1997) and uses an auditory-inspired modulation filterbank to group modulation frequency bins; with the Ratio_2000 measure, simple averaging of Fourier transform-derived frequency bins is performed.

### 2.4. Nasality

Previous research has suggested that formant frequencies and their bandwidths can be used to characterize nasality. As examples, the bandwidth of the first formant frequency was shown to be related to nasality (Fant, 1960; Baken and Orlikoff, 2000; O'Shaughnessy, 2008); formant shifts were observed in (House and Stevens, 1956; Baken and Orlikoff, 2000); and more recently, the first four formants and their bandwidths were used to classify nasality (Zecevic, 2002).

We investigate the use of the first two formant frequencies ($F_1$ and $F_2$) and their bandwidths ($BW_1$ and $BW_2$) for dysarthric word intelligibility estimation. The open source Wavesurfer software (Sjolander and Beskow, 2000) was used to compute $F_i$ and $BW_i$ ($i = 1$ and 2), over voiced speech segments, using the tracking algorithm described in (Talkin, 1987). Default parameters were used, more specifically: 12th order LP analysis, 0.7 pre-emphasis factor, and 49 ms (Hamming) analysis windows with a 10 ms window shift. Here, the average and standard deviation of the per-frame estimated formant frequencies ($\overline{F}_i$ and $\sigma_{F_i}$, respectively) and bandwidths ($\overline{BW}_i$ and $\sigma_{BW_i}$) are tested as correlates of word intelligibility.

### 2.5. Prosody

Three conventional parameters are computed in order to characterize disordered prosody (Bunton et al., 2000; Schlenck et al., 1993). The first two parameters are related to the variation in fundamental frequency ($f0$), namely, $f0$ standard deviation ($\sigma_{f0}$) and $f0$ range ($\Delta_{f0}$) (see Schlenck et al. (1993)). Pitch estimates are computed using a robust adaptive pitch tracker algorithm (Talkin, 1995). The third parameter corresponds to the voicing percentage, computed as the ratio between the duration of voiced segments in the uttered word and the duration of the entire utterance. The measure is represented by $\%\mathcal{V}$ and has been used in the past to characterize speech disorders (Maier et al., 2009; Colcord and Adams, 1979).

### 2.6. Composite measure

Previous subjective perceptual studies of dysarthric speech suggest that intelligibility can be expressed as a weighted linear combination of features from different perceptual dimensions, such as articulation, vocal quality, nasality, and prosody (De Bodt et al., 2002). Hence, in order to further improve intelligibility estimation performance, a composite measure is developed below consisting of a linear combination of the six key parameters shown to correlate significantly with subjective scores (see parameters with correlation values in bold in Table 2). Similar to (De Bodt et al., 2002), nasality was shown to contribute marginally to intelligibility estimation, thus formant-related parameters are not included in the composite measure given by

$$f = A_0 + A_1 * \mathcal{K}_{LP} + A_2 * \sigma_\Delta + A_3 * \mathrm{LHMR} \\ + A_4 * \sigma_{f0} + A_5 * \Delta_{f0} + A_6 * \%\mathcal{V}.$$

In order to gauge the degree of influence of each constituent measure, the six parameters are normalized to zero mean and unit variance. To estimate the weights $A_i$, the available data is partitioned into mutually exclusive training and test datasets. The weights obtained using the (normalized) training data subsets are described in Section 3.3.

## 3. Experimental setup

### 3.1. Database: UA-speech

The data used in our experiments consisted of the audio content of the Universal Access (UA-Speech) audio-visual database made publicly-available by the University of Illinois (Kim et al., 2008). Disordered speech data was collected using an eight-microphone array, sampled at 16 kHz and digitized with 16-bit precision. In the array, adjacent microphones were spaced apart by 1.5 inches and each microphone sized 6 mm in diameter. One channel of the array was reserved for recording DTMF tones, which served as flags for subsequent offline word segmentation. In this experiment, speech recorded from microphone no. 6 of the array was used.

Seventeen participants diagnosed with cerebral palsy were recruited from the Rehabilitation Education Center at the University of Illinois at Urbana-Champaign and from the Trace Research and Development Center at the University of Wisconsin-Madison. Participants were seated comfortably in front of a laptop computer and asked to read an isolated word displayed on a computer screen. Each participant read three blocks of words, totalling 765 isolated word utterances per participant. Each block contained 255 words, including 155 words that were repeated in each block and 100 uncommon words that differed across blocks. The repeated words consisted of the 10 digits (‘zero’ to ‘nine’), 26 radio alphabet letters (e.g., ‘Alpha’, ‘Bravo’), 19 computer commands (e.g., ‘backspace’, ‘delete’), and the 100 most common words in the Brown corpus of written English (e.g., ‘it’, ‘is’, ‘you’). The 300 uncommon words (100 per block) were selected from children’s novels digitized by Project Gutenberg and consisted of words such as ‘naturalization’ and ‘moonshine’ (Kim et al., 2008).

To assess speech intelligibility, a subjective listening test was performed. Two hundred words were selected for the test, including 10 digits, 25 radio alphabet letters, 19 computer commands, and 73 words randomly selected from each of the common and uncommon word categories. For the purpose of intra-listener reliability assessment, 25 words out of the 200 were arbitrarily chosen and uttered twice in the list. The final 225 speech files were randomly ordered (with a constraint that repeated words were not adjacent to each other) and presented to the listeners. Five naive listeners were recruited for each speaker; listeners were between the ages of 18–40, native speakers of American English, had no prior experience with disordered speech, and had no previous training in phonetic transcription.

Listeners were instructed to provide orthographic transcriptions of each of the 225 speech utterances presented via headphones in a quiet environment; they were allowed to listen to the words as many times as needed. Listener transcriptions were then analyzed and the mean percentage of correct responses, averaged across the five listeners, was calculated to obtain the subjective intelligibility score of each dysarthric speaker; an average intra-listener agreement rate of 91.64% was obtained. Based on the averaged intelligibility score, each speaker was then classified into one of four intelligibility categories, namely,: very low (0–25%), low (26–50%), mid (51–75%) and high (76–100%). Below, data from the ten participants with spastic dysarthria were used; Table 1 summarizes their demographics.

### 3.2. Figures of merit

In order to assess the performance of the proposed and benchmark prosody measures, the Pearson correlation coefficient ($R$) is used and given by

$$R = \frac{\sum_{i=1}^{N}(w_i - \bar{w})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(w_i - \bar{w})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}, \tag{15}$$

where $w_i$ corresponds to the $i$th participant’s subjective intelligibility score, $y_i$ is the value of a measure calculated over the speech data of participant $i$, $\bar{w}$ the average of $w_i$, and $\bar{y}$ the average of $y_i$. Ultimately, the aim in objective intelligibility estimation is to develop a measure that ranks similarly with the subjective listening ratings. As a consequence, the Spearman rank correlation coefficient ($R_S$) is also used as a figure of merit. Spearman correlation is computed in a manner similar to (15), except that the intelligibility and correlate values

Table 1
Demographics of the ten spastic dysarthric speakers.

| Subject | Gender | Age | Intelligibility (%) | Category |
|---|---|---|---|---|
| 1 | Male | 18 | 2 | Very low |
| 2 | Male | 18 | 15 | Very low |
| 3 | Male | 58 | 28 | Low |
| 4 | Male | Unreported | 43 | Low |
| 5 | Male | 21 | 58 | Mid |
| 6 | Male | 40 | 91 | High |
| 7 | Male | 28 | 93 | High |
| 8 | Female | 51 | 6 | Very low |
| 9 | Female | 30 | 29 | Low |
| 10 | Female | 22 | 95 | High |

Please cite this article in press as: Falk, T.H. et al., Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, Speech Comm. (2011), doi:10.1016/j.specom.2011.03.007

are replaced by their rank values. Lastly, root-mean-square error ($RMSE$) is used to assess word intelligibility measurement accuracy of the proposed composite measures; $RMSE$ is computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(w_i - y_i)^2}{N}}, \tag{16}$$

### 3.3. Experimental results

Table 2 reports the Pearson ($R$) and Spearman rank ($R_S$) correlation coefficients, along with their corresponding $p -$ values (not to be confused with the order $p$ of the LP analysis), for the afore-developed measures. As can be seen, the proposed $\mathcal{K}_{LP}$ measure achieved significant correlations ($p < 0.05$) with subjective intelligibility scores. Of the four proposed temporal dynamics measures, two were shown to correlate significantly with subjective ratings; one characterized short-term ($\sigma_\Delta$) and the other long-term (LHMR) temporal dynamics perturbations. Additionally, formant-related features did not correlate significantly with listener scores, whereas all three prosodic parameters were shown to correlate significantly (negatively).

As mentioned in Section 2.6, in order to develop the composite measure $f$, the UA-Speech database was partitioned into two disjoint sets. Speech files belonging to the 'uncommon word' category (300 files per participant) served as (unseen) test data and the remaining files (465 files per participant) served as training data and were used to obtain the weights $A_i$ in (14). The last four rows of Table 2 report the performance of the composite measure on the test set under four different scenarios. The first ($f_{raw}$) reports the performance of the "raw" intelligibility estimates obtained directly from (14). The results for $f_{map}$ are obtained after a third-order monotonic polynomial regression is applied to the raw scores. This type of mapping is commonly used in objective voice quality estimation tasks and maps the objective scores onto the subjective scale (ITU-T P.563, 2004; Falk and Chan, 2006). The mapping provides a scale adjustment but does not alter the ranking of the objective scores, as can be seen by the $R_S$ values reported in Table 2. The plots in Fig. 4 depict the subjective *versus* estimated intelligibility scores before and after the 3rd order mapping. As can be seen, the mapping adjusts the estimates to better represent the subjective scale.

As reported by Schlenck et al. (1993), speech impairments may differ not only with dysarthria type, but also by the severity of the disorder. Accordingly, we explore further partitioning the training and test sets into mid-low (0–50%) and mid-high (51–100%) intelligibility classes. These two classes are chosen as they can be easily categorized by even an unexperienced therapist. The sub-partitions are then used to train two "class-based" linear estimators ($f_{class}$) where the subscript *class* corresponds to either the 'mid-low' or 'mid-high' intelligibility classes. The overall performance obtained with the class-based estimators, both before ($f_{class}$) and after a 3rd order monotonic polynomial mapping ($f_{class,map}$), are reported in the last two rows of Table 2 and are shown in Fig. 4. As observed, the class-based estimators further improve intelligibility estimation and result in significant correlations with subjective scores and near-perfect rank correlation ($R_S = 0.96$). Moreover, the four composite measures, $f_{raw}$, $f_{map}$, $f_{class}$, and $f_{class,map}$ achieved RMSE values of 18.6, 10.4, 10.2 and 8.6, respectively. Lastly, Table 3 reports the obtained $A_i$ weights using the different composite

Table 2
Performance comparison for the proposed and benchmark prosody measures. Performances of the composite measure are reported before ($f_{raw}$) and after ($f_{map}$) a 3rd order monotonic polynomial mapping, as well as with severity classification, both before ($f_{class}$) and after ($f_{class,map}$) the 3rd order polynomial mapping.

| Measure | $R$ | $p$ | $R_S$ | $p$ |
|---|---|---|---|---|
| *A typical vocal source excitation* | | | | |
| $\mathcal{K}_{LP}$ | **0.88** | 0.001 | **0.81** | 0.005 |
| *Temporal dynamics* | | | | |
| $\sigma_\Delta$ | **0.71** | 0.020 | **0.81** | 0.005 |
| $\mathcal{S}_\Delta$ | 0.44 | 0.200 | 0.42 | 0.230 |
| $\mathcal{K}_\Delta$ | −0.16 | 0.650 | −0.01 | 0.980 |
| LHMR | **−0.65** | 0.040 | **−0.70** | 0.020 |
| *Nasality* | | | | |
| $\overline{F}_1$ | 0.46 | 0.070 | 0.50 | 0.060 |
| $\overline{F}_2$ | 0.21 | 0.560 | 0.19 | 0.600 |
| $\sigma_{F_1}$ | 0.12 | 0.740 | 0.19 | 0.600 |
| $\sigma_{F_2}$ | 0.08 | 0.830 | 0.10 | 0.800 |
| $\overline{BW}_1$ | −0.44 | 0.070 | −0.46 | 0.063 |
| $\overline{BW}_2$ | −0.05 | 0.740 | −0.15 | 0.600 |
| $\sigma_{BW_1}$ | −0.10 | 0.650 | −0.19 | 0.600 |
| $\sigma_{BW_2}$ | −0.13 | 0.320 | −0.36 | 0.470 |
| *Prosody* | | | | |
| $\sigma_{f0}$ | **−0.57** | 0.050 | −0.43 | 0.210 |
| $\Delta_{f0}$ | **−0.72** | 0.010 | **−0.76** | 0.010 |
| $\%\mathcal{V}$ | **−0.77** | 0.009 | **−0.75** | 0.001 |
| *Composite* | | | | |
| $f_{raw}$ | 0.94 | $10^{-5}$ | 0.89 | $10^{-4}$ |
| $f_{map}$ | 0.95 | $10^{-5}$ | 0.89 | $10^{-4}$ |
| $f_{class}$ | 0.96 | $10^{-5}$ | 0.96 | $10^{-6}$ |
| $f_{class,map}$ | 0.97 | $10^{-5}$ | 0.96 | $10^{-6}$ |

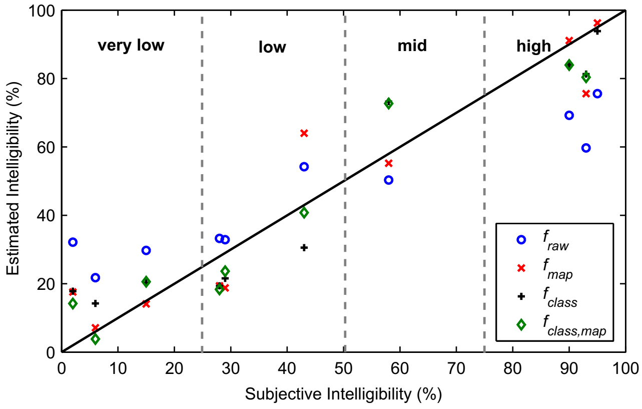

Fig. 4. Estimated *versus* subjective intelligibility obtained with the proposed composite measure under four scenarios, namely, $f_{raw}$, $f_{map}$, $f_{class}$ and $f_{class,map}$.

Please cite this article in press as: Falk, T.H. et al., Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, Speech Comm. (2011), doi:10.1016/j.specom.2011.03.007

Table 3
Weights $A_i$ of the composite measures $f_{raw}$ and $f_{class}$, where *class* assumes labels 'mid-low' and 'mid-high'.

| Weight | $f_{raw}$ | $f_{mid-low}$ | $f_{mid-high}$ |
|---|---|---|---|
| $A_0$ | 45.82 | 24.99 | 77.53 |
| $A_1$ | 13.68 | 5.81 | 2.92 |
| $A_2$ | 8.92 | 6.59 | 4.27 |
| $A_3$ | −0.87 | −0.10 | 1.72 |
| $A_4$ | 1.34 | 0.79 | 9.08 |
| $A_5$ | −8.83 | −1.58 | −8.73 |
| $A_6$ | −6.36 | −0.98 | −7.58 |

measurement configurations, namely $f_{raw}$ and $f_{class}$ for classes mid-low and mid-high.

## 4. Discussion

This paper has described several objective measures of different perceptual dimensions, such as vocal quality, articulation, nasality, and prosody. It was shown that when a salient subset of the measures are linearly combined, the resultant composite measure outperforms the constituent measures individually. This finding resonates closely with those reported by De Bodt et al. (2002) for subjective perceptual tests. As an example, on the test dataset used to obtain the results reported for the composite measure in Table 2, the best individual feature, namely $\mathcal{K}_{LP}$, achieved $R = 0.88$ ($p = 0.001$) and $R_S = 0.81$ ($p = 0.005$) when used alone. As a consequence, the proposed $f_{class}$ composite measure attains an approximate 67% correlation-improvement ($R$%) relative to using the best single measure alone. Here, ($R$%) is computed as

$$R\% = \frac{R_{composite} - R_{individual}}{1 - R_{individual}} \times 100\% \tag{17}$$

and reflects the percentage reduction of the individual measure's performance gap to perfect correlation (Falk and Chan, 2006).

Commonly, dysarthric speech is considered monotone and "robotic" (Klopfenstein, 2009). Hence, it is expected that for more severe cases of dysarthria (low intelligibility) lower pitch variability/range is obtained. The negative correlations reported in Table 2 for prosody-related features, however, suggest otherwise. While these findings may seem counterintuitive, they are inline with those reported by Schlenck et al. (1993), where the nature of dysprosody was shown to vary with the severity of dysarthria. More specifically, monotonicity was reported for mild dysarthric speakers only and higher pitch variation/range was observed for speakers with severe disorders. This same behaviour was observed with the prosody-related measures used here. As examples, average $\sigma_{f0} = 24$ Hz and $\Delta_{f0} = 92$ Hz were observed for mid-high intelligibility level speakers, whereas $\sigma_{f0} = 40$ Hz and $\Delta_{f0} = 165$ Hz were found for mid-low intelligibility speakers.

Careful scrutiny of the weights reported in Table 3 suggest that other severity class dependencies may also exist. For example, short-term temporal dynamics and vocal harshness, represented by parameters $\sigma_\Delta$ and $\mathcal{K}_{LP}$ respectively, have greater influence on intelligibility prediction for the mid-low class than for the mid-high class. Prosody-related parameters, on the other hand, have greater influence on prediction for the mid-high class, corroborating findings reported by Schlenck et al. (1993). These findings suggest that composite measures should be trained separately for different degrees of dysarthria, as was the case with the proposed $f_{class}$ measure.

Severity class dependence was also observed for the proposed LHMR measure. As seen from the weight $A_3$ in Table 3, a negative weight was obtained for mid-low speakers, whereas a positive weight was found for mid-high intelligibility speakers. This finding resonates closely with our hypothesis that the unclear distinction between adjacent phonemes will cause a shift of the significant modulation frequency content to modulation frequencies below 4 Hz. For severe disorders, higher LHMR values (i.e., greater modulation spectral frequency content below 4 Hz) should cause intelligibility levels to decrease. With mild dysarthria, on the other hand, the significant modulation frequency content is better spread across higher modulation frequencies, much like what has been observed for "healthy" speech (Drullman et al., 1994a). With healthy speech, however, modulation spectral content greater than 20 Hz is often associated with unnatural speech components (e.g., noise) (Kim, 2004). As a consequence, higher LHMR values (i.e., lower modulation spectral content beyond 4 Hz) should cause intelligibility levels to increase and this is reflected by the positive $A_3$ weight found for mid-high intelligibility speakers (see Table 3).

Notwithstanding, for both the mid-low and mid-high classes, the LHMR parameter has the lowest influence on overall intelligibility prediction. This may be due to the fact that the UA-Speech corpus used here is comprised of single-word utterances. The long-term temporal disturbances captured by the LHMR parameter are likely to result in greater prediction influence when used with longer-duration utterances (e.g., sentences, such as in (LeGendre et al., 2009)) or with running speech.

Similar to the results reported in (De Bodt et al., 2002), nasality was shown to be the least important dimension in word intelligibility estimation and the investigated formant-related features did not correlate significantly with subjective listener scores. Notwithstanding, we investigated the inclusion of parameters $\overline{F}_1$ and $\overline{BW}_1$ in (14) but did not observe any improvement in intelligibility estimation performance. While nasality may be prominent in spastic dysarthric speech, it does not seem to affect word intelligibility prediction.

The developed measures have only been tested on adult speakers with *spastic* dysarthria associated with cerebral palsy. Future investigations should also focus on other types of dysarthria (e.g., ataxic, hypokinetic), on children's speech (Saz et al., 2008), as well as on a more gender-balanced participant pool, as gender differences may also

Please cite this article in press as: Falk, T.H. et al., Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, Speech Comm. (2011), doi:10.1016/j.specom.2011.03.007

play a factor in intelligibility estimation (Schlenck et al., 1993). Additionally, objective intelligibility measures may not only assist clinicians in treatment evaluation, but may also improve the performance of assistive technologies based on automatic speech recognition (e.g., Hasegawa-Johnson et al., 2006). These technologies may improve the communication ability of individuals with speech disorders, thus ultimately enhancing their quality of life.

## 5. Conclusion

In this paper, several measures were proposed to characterize vocal source excitation oddity and temporal dynamics perturbations and shown to correlate significantly with subjective word intelligibility ratings. A composite measure was also developed based on linearly combining a salient subset of the proposed measures and conventional prosody-related measures; the composite measure was shown to be a reliable indicator of dysarthric word intelligibility. To assist clinicians in the diagnosis and treatment of speech disorders, the composite measure can also be employed, in conjunction with the constituent parameters of the measure, to characterize individual intelligibility degradation factors such as vocal harshness, articulation, and dysprosody.

## Acknowledgements

The authors wish to acknowledge Dr. Mark Hasegawa-Johnson for making the UA-Speech database available, the Natural Sciences and Engineering Research Council of Canada for their financial support, and the anonymous reviewers for their insightful comments.

## References

Ananthapadmanabha, T., Yegnanarayana, B., 1979. Epoch extraction from linear prediction residual for identification of closed glottis interval. IEEE Trans. Acoust. Speech Signal Process. 27 (4), 309–319.

Arai, T., Pavel, M., Hermansky, H., Avendano, C., 1996. Intelligibility of speech with filtered time trajectories of spectral envelopes. In: Proc. Internat. Conf. on Speech and Language Processing, pp. 2490–2493.

Baken, R., Orlikoff, R., 2000. Clinical Measurement of Speech and Voice. Singular Publishing Group.

Benesty, J., Chen, J., Huang, Y., 2008. Linear Prediction. Springer Handbook of Speech Processing. Springer-Verlag, Berlin.

Bunton, K., Kent, R., Kent, J., Rosenbek, J., 2000. Perceptuo-acoustic assessment of prosodic impairment in dysarthria. Clin. Linguist. Phonet. 14 (1), 13–24.

Colcord, R., Adams, M., 1979. Voicing duration and vocal SPL changes associated with stuttering reduction during singing. J. Speech Hear. Res. 22 (3), 468.

Constantinescu, G., Theodoros, D., Russell, T., Ward, E., Wilson, S., Wootton, R., 2010. Assessing disordered speech and voice in Parkinson's disease: A telerehabilitation application. Internat. J. Language Comm. Disord. 45 (6), 630–644.

Dau, T., Kollmeier, B., Kohlrausch, A., 1997. Modeling auditory processing of amplitude modulation. I – Modulation detection and masking with narrowband carriers. J. Acoust. Soc. Amer. 102, 2892–2905.

Dau, T., Puschel, D., Kohlrausch, A., 1996. A quantitative model of the effective signal processing in the auditory system. I – model structure. J. Acoust. Soc. Amer. 99 (6), 3615–3622.

De Bodt, M., Hernández-Dı́az Huici, M., Van De Heyning, P., 2002. Intelligibility as a linear combination of dimensions in dysarthric speech. J. Comm. Disord. 35 (3), 283–292.

Doyle, P., Leeper, H., Kotler, A., Thomas-Stonell, N., O'Neill, C., Dylke, M., Rolls, K., 1997. Dysarthric speech: A comparison of computerized speech recognition and listener intelligibility. J. Rehabil. Res. Dev. 34 (3), 309–316.

Drullman, R., Festen, J., Plomp, R., 1994a. Effect of reducing slow temporal modulations on speech reception. J. Acoust. Soc. Amer. 95 (5), 2670–2680.

Drullman, R., Festen, J., Plomp, R., 1994b. Effect of temporal envelope smearing on speech reception. J. Acoust. Soc. Amer. 95 (2), 1053–1064.

Duffy, J., 2005. Motor Speech Disorders: Substrates. In: Differential Diagnosis, and Management. Mosby, St. Louis.

Falk, T., Chan, W.-Y., 2010. Temporal dynamics for blind measurement of room acoustical parameters. IEEE Trans. Instrum. Measure. 59 (4), 978–989.

Falk, T., Zheng, C., Chan, W., 2010. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. IEEE Trans. Audio Speech Lang. Process. 18 (7), 1766–1774.

Falk, T.H., Chan, W.-Y., 2006. Single-ended speech quality measurement using machine learning methods. IEEE Trans. Audio Speech Lang. Process. 14 (6), 1935–1947.

Fant, G., 1960. Nasal Sounds and Nasalization. Acoustic Theory of Speech Production. Mouton, The Hague, Netherlands.

Ferrier, L., Shane, H., Ballard, H., Carpenter, T., Benoit, A., 1995. Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. Augmentative Altern. Comm. 11 (3), 165–175.

Gillespie, B., Malvar, H., Florêncio, D., 2001. Speech dereverberation via maximum-kurtosis subband adaptive filtering. In: Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing, vol. 6.

Glasberg, B., Moore, B., 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47 (1–2), 103–138.

Green, P., Carmichael, J., Hatzis, A., Enderby, P., Hawley, M., Parker, M., 2003. Automatic speech recognition with sparse training data for dysarthric speakers. In: Proc. Eighth European Conf. on Speech Communication and Technology, p. 4.

Gu, L., Harris, J., Shrivastav, R., Sapienza, C., 2005. Disordered speech assessment using automatic methods based on quantitative measures. EURASIP J. Appl. Signal Process. 9, 1400–1409.

Hasegawa-Johnson, M., Gunderson, J., Perlman, A., Huang, T., 2006. HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria. In: Proc. Internat. Conf. on Acoustics, Speech, and Signal Processing, pp. 1060–1063.

Hill, A., Theodoros, D., Russell, T., Cahill, L., Ward, E., Clark, K., 2006. An Internet-based telerehabilitation system for the assessment of motor speech disorders: A pilot study. Amer. J. Speech-Lang. Pathol. 15 (1), 45.

House, A., Stevens, K., 1956. Analog studies of the nasalization of vowels. J. Speech Hear. Disord. 21 (2), 218.

Huang, X., Acero, A., Hon, H.-W., 2001. Spoken Language Processing: A Guide to Theory. In: Algorithm and System Development. Prentice-Hall, New Jersey.

ITU-T P.563, 2004. Single-ended method for objective speech quality assessment in narrowband telephony applications, Geneva, Switzerland.

Kent, R., Kent, J., Weismer, G., Martin, R., Sufit, R., Brooks, B., Rosenbek, J., 1989. Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects. Clin. Linguist. Phonet. 3 (4), 347–358.

Kim, D.-S., 2004. A cue for objective speech quality estimation in temporal envelope representation. IEEE Signal Process. Lett. 11 (10), 849–852.

Please cite this article in press as: Falk, T.H. et al., Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, Speech Comm. (2011), doi:10.1016/j.specom.2011.03.007

Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J., Huang, T., Watkin, K., Frame, S., 2008. Dysarthric speech database for universal access research. In: Proc. Internat. Conf. on Spoken Language Processing, pp. 1741–1744.

Klopfenstein, M., 2009. Interaction between prosody and intelligibility. Internat. J. Speech-Lang. Pathol. 11 (4), 326–331.

LeGendre, S., Liss, J., Lotto, A., 2009. Discriminating dysarthria type and predicting intelligibility from amplitude modulation spectra. J. Acoust. Soc. Amer. 125, 2530–2531.

Maier, A., Haderlein, T., Eysholdt, U., Rosanowski, F., Batliner, A., Schuster, M., Noth, E., 2009. PEAKS-A system for the automatic evaluation of voice and speech disorders. Speech Comm. 51 (5), 425–437.

Middag, C., Martens, J.-P., Van Nuffelen, G., De Bodt, M., 2009. Automated intelligibility assessment of pathological speech using phonological features. EURASIP J. Adv. Signal Process., 9. Article ID: 629030.

O'Shaughnessy, D., 2008. Formant Estimation and Tracking. Springer Handbook of Speech Processing. Springer-Verlag, Berlin, pp. 213–227.

Picone, J., 1993. Signal modeling techniques in speech recognition. Proc. IEEE 81 (9), 1215–1247.

Raghavendra, P., Rosengren, E., Hunnicutt, S., 2001. An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition systems. Augmentative Altern. Comm. 17 (4), 265–275.

Rudzicz, F., 2007. Comparing speaker-dependent and speaker-adaptive acoustic models for recognizing dysarthric speech. Proc. Internat. ACM SIGACCESS Conf. on Computers and Accessibility. ACM, pp. 256–257.

Saz, O., Rodriguez, W., Lleida, E., Vaquero, C., 2008. A novel corpus of children's disordered speech. In: Proc. First Workshop on Child, Computer and Interaction, p. 6.

Schlenck, K., Bettrich, R., Willmes, K., 1993. Aspects of disturbed prosody in dysarthria. Clin. Linguist. Phonet. 7 (2), 119–128.

Sharma, H., Hasegawa-Johnson, M., Gunderson, J., Perlman, A., 2009. Universal access: Preliminary experiments in dysarthric speech recognition. In: Proc. 10th Annual Conf. of the Internat. Speech Communication Association, p. 4.

Sjolander, K., Beskow, J., 2000. Wavesurfer-an open source speech tool. In: Proc. Internat. Conf. on Spoken Language Processing, p. 4.

Slaney, M., 1993. An efficient implementation of the Patterson–Holdsworth auditory filterbank. Tech. rep., Apple Computer, Perception Group.

Talkin, D., 1987. Speech formant trajectory estimation using dynamic programming with modulated transition costs. J. Acoust. Soc. Amer. 82, S55.

Talkin, D., 1995. A Robust Algorithm for Pitch Tracking (RAPT). Speech Coding and Synthesis. Elsevier Science Publishers, Amsterdam, pp. 495–518.

Van Nuffelen, G., Middag, C., De Bodt, M., Martens, J., 2009. Speech technology-based assessment of phoneme intelligibility in dysarthria. Internat. J. Lang. Comm. Disord. 44 (5), 716–730.

Zecevic, A., 2002. Ein sprachgestutztes trainingssystem zur evaluierung der nasalitiat. Ph.D. thesis, Universitat Mannheim.

Please cite this article in press as: Falk, T.H. et al., Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, Speech Comm. (2011), doi:10.1016/j.specom.2011.03.007