Article:

Alhinti, L., Christensen, H. orcid.org/0000-0003-3028-5062 and Cunningham, S. orcid.org/0000-0001-9418-8726 (2021) Acoustic differences in emotional speech of people with dysarthria. Speech Communication, 126. pp. 44-60. ISSN 0167-6393

https://doi.org/10.1016/j.specom.2020.11.005

Speech Communication 126 (2021) 44–60

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Acoustic differences in emotional speech of people with dysarthria

Lubna Alhinti [a,*], Heidi Christensen [a,c], Stuart Cunningham [b,c]

[a] Department of Computer Science, University of Sheffield, United Kingdom
[b] Health Science School, University of Sheffield, United Kingdom
[c] Centre of Assistive Technology and Connected Healthcare (CATCH), Sheffield, United Kingdom

A R T I C L E I N F O

Keywords:
Acoustic analysis
Dysarthria
Paralinguistic
Emotion expression

A B S T R A C T

Communicating emotion is essential in building and maintaining relationships. We communicate our emotional state not just with the words we use, but also how we say them. Changes in the rate of speech, short-term energy and intonation all help to convey emotional states like 'angry', 'sad' and 'happy'. People with dysarthria, the most common speech disorder, have reduced articulatory and phonatory control. This can affect the intelligibility of their speech, especially when communicating with unfamiliar conversation partners. However, we know little about how people with dysarthria convey their emotional state, and whether they are having to make changes to their speech to achieve this. In this study, we investigated the ability of people with dysarthria, caused by cerebral palsy and Parkinson's disease, to communicate emotions in their speech, and we compared their speech to that of speakers with typical speech. A parallel database of emotional speech was collected. One female speaker with dysarthria due to cerebral palsy, 3 speakers with dysarthria due to Parkinson's disease (2 female and 1 male), and 21 typical speakers (9 female and 12 male) produced sentences with 'angry', 'happy', 'sad', and 'neutral' emotions. A number of acoustic features were analysed using linear multi-level modeling. The results show that people with dysarthria were able to control some aspects of the suprasegmental and prosodic features when attempting to communicate emotions. For most speakers the changes they made are consistent with the changes made by speakers with typical speech. Even when the changes might be different to that of typical speakers, acoustic analysis shows these were consistent for different emotions. The analysis shows that variation in energy and jitter (local absolute) are major indicators of emotion in the study.

1. Introduction

People with speech disorders, such as dysarthria, can find it difficult to communicate with unfamiliar conversation partners. There is evidence that people can quickly adapt to speech from an unfamiliar person with a speech disorder, but often people can still find such speech difficult to understand. We know that some of this difficulty comes from the fact that people with a speech disorder are sometimes unable to produce speech sounds accurately, or use typical intonation. However, we do not know if paralinguistic information, such as the emotional state of the speaker, which can assist a listener, is conveyed by people with disorders like dysarthria.

Dysarthria is a neurological disorder that affects different aspects of speech production and is the most common motor speech disorder (Walshe and Miller, 2011). It can cause weakness in the muscles responsible for speaking, miscoordination or inaccurate articulatory movements, and irregularity in the tone, steadiness, or speed (Duffy, 2013). Dysarthric speech has been characterized prosodically as having monoloudness, monopitch, impaired ranges of F0, vocal intensity, and rate. Darley et al. (1969) outlined five different types of dysarthria each of which sound different depending on the associated speech and voice dimensions.

People with dysarthria can struggle to be understood in conversation, not only because the intelligibility of their words is affected, but also because their paralinguistic information can be limited. Generally, human-to-human communication can be viewed as the process of producing and receiving messages. The messages themselves are formulated using different signs and codes that are interpreted by the receiver (Steinberg, 1995). However, people do not express all their feelings through words; nonverbal information conveys part of a person's feelings and emotions that affects the meaning of the spoken word (Calero, 2005). Thus, relying on verbal communication only may lead to problems in communication and may increase the potential of being socially withdrawn. This paper is the first study of its kind to investigate to what

* Corresponding author.
E-mail address: laalhinti1@sheffield.ac.uk (L. Alhinti).

https://doi.org/10.1016/j.specom.2020.11.005
Received 16 January 2020; Received in revised form 10 October 2020; Accepted 26 November 2020
Available online 4 December 2020

extent people with dysarthria are able to communicate emotions and how their acoustic *signaling* of emotions might differ to that used by typical speakers.

It has been shown that emotions have a direct effect on vocal production (Scherer et al., 1980). Different emotions are usually indexed by specific acoustic characteristics. Physiological changes that result from being in a particular emotional state affect the phonation, respiration, and articulation in a way that creates specific acoustic characteristics for each emotion (Scherer, 2003; Banse and Scherer, 1996). Although the exact details of which acoustic parameters are affected and what changes on these acoustic parameters occur while being in a certain emotion are still not very clear, there are however, some features that usually change in emotional speech such as the fundamental frequency (F0) and the energy (Banse and Scherer, 1996). The analysis of the F0 (minimum, maximum, range, mean, etc.) has been included in most of the research on emotional speech. F0 statistics are one of the features that correlate with emotional vocal expressions. Higher F0 is usually associated with high-arousal emotions such as ‘angry’ and ‘happy’, while lower F0 is more associated with low-arousal emotions such as ‘sad’ (Breitenstein et al., 2001; Johnstone and Scherer, 2000; Guo et al., 2016). Also, ‘happy’ and ‘angry’ are found to be associated with a very wide range of F0 values compared to neutral speech while ‘sad’ is found to be associated with a less wide range of F0 values (Guo et al., 2016; Murray and Arnott, 1993). Research has also shown that higher energy is usually associated with high-arousal emotions such as ‘angry’ and ‘happy’, while lower energy is more associated with low-arousal emotions such as ‘sad’ (Johnstone and Scherer, 2000). Scherer (2003) and Johnstone and Scherer (2000) have presented more detailed results on the effect of different emotions on selected acoustic parameters. All of these studies were conducted on typical speech.

The literature includes a number of studies that compares acoustic differences between dysarthric and typical speech (Darley et al., 1975; Platt et al., 1980; Le Dorze et al., 1994; Kempler and Van Lancker, 2002; Bunton and Weismer, 2001; Ansel and Kent, 1992; Rosen et al., 2006; Connaghan and Patel, 2017; Liss and Weismer, 1994; Patel, 2004). Despite having speech which is less intelligible, many studies show that even with the limited phonological and prosodic dimensions, many people with dysarthria have enough control to signal prosodic contrast on different tasks. For example, several studies investigated the ability of people with dysarthria caused by either cerebral palsy or Parkinson’s disease to signal question-statement contrast in different languages (Patel, 2002b, 2003; Liu et al., 2019; Ma and Hoffmann, 2010; Ma et al., 2010). Other studies investigated the prosodic and acoustic characteristics such as F0, intensity, and speech rate of speakers with dysarthria in comparison to typical speakers (Patel, 2002a; Canter, 1963; Illes et al., 1988; Gu et al., 2017; Hammen and Yorkston, 1996; Ghio et al., 2014; Rusz et al., 2011; (Holmers et al., 2000). It is important to note that some of these studies were carried out on small samples, in different languages that have different characteristics, and/or on speakers with different severity levels of dysarthria. Thus, any attempt of generalizing the findings is difficult, however, there are several consistent points. First, the majority of these studies show that even with having different prosodic characteristics to typical speakers, speakers with dysarthria were able to signal the question-statement contrast. Second, high inter-speaker variability among speakers with dysarthria was observed in some of these studies. Third, there is a lack of consistency in some of the reported results on the prosodic and acoustic characteristics between speakers with dysarthria and typical speakers. For example, some studies reported higher average F0 for speakers with dysarthria compared to typical speakers while others reported no differences between the two groups of speakers. Nevertheless, these studies show that people with dysarthria may have enough control of prosodic and phonatory features that allows them to communicate emotions, convey intentions, and obtain listeners’ attention. This potential control opens many new doors of investigations related to the kinds of paralinguistic information people with dysarthria can communicate and their consistency of doing so, i.e., the intra- and inter-speakers variability when communicating specific information.

To better understand communication of emotion by people with dysarthria, the authors conducted a survey (Alhinti et al., 2020). The survey addressed several points including how difficult it is for people with dysarthria to communicate emotions, and what emotions are important for them to be able to communicate successfully. The survey was answered by eight participants with moderate and severe levels of dysarthria caused by cerebral palsy. The results indicated that being able to successfully communicate ‘happiness’ is very important for them in social and in everyday settings. ‘Anger’ and ‘sadness’ also appeared to be among the most important emotions that people with dysarthria would like to be able to communicate successfully. In regards to what emotion is difficult for them to get across, ‘anger’ was indicated by almost half of the respondents. Although their speech is often perceived as being unintelligible, people with dysarthria show a strong preference for using it as their way of communication (Beukelman et al., 2007). This is expected, as speech is the natural and fastest way of communication.

Many studies have focused on understanding the articulation errors in dysarthric speech and finding ways to automatically recognize their speech. In addition, as mentioned above, a number of studies have investigated the prosodic control ability of speakers with dysarthria in different tasks. However, their ability to convey emotions in their speech through suprasegmental and prosodic features remains unexplored. This work is motivated by a long term goal to improve voice-input communication aids used by people with dysarthria in a way that makes it more sensitive to specific cues in the vocalization signal produced by the speaker with dysarthria and hence act more according to the speaker’s intention. This study investigates which acoustic characteristics people with dysarthria use to signal the different emotions, and if these are different to typical speakers.

The purpose of this study is to answer the following questions. First, can people with dysarthria due to cerebral palsy or Parkinson’s disease make systematic changes to their speech to convey their emotional state? Second, if they are able to make such changes are these similar to those made by speakers with typical speech?

2. Method

The study was approved by the ethical review panel of the Department of Computer Science at the University of Sheffield.

2.1. Dataset

Since there are no available emotional databases on dysarthric speech, we collected a parallel database of typical and dysarthric emotional speech. The fact that emotional states are caused by many factors is the reason behind the difficulty of collecting samples of people under a particular emotional state (Douglas-Cowie et al., 2003). Over the past decades, there has been considerable debate over the type of methodology that should be used to collect emotional databases. The three methodologies that have been used for databases of typical emotional speech are: natural, elicited (or induced), and acted. Each methodology has its advantages and disadvantages. Adopting the natural methodology, i.e., recording spontaneous emotions resulting from natural stimulus such as recordings of pilots in a dangerous flight situations, was not appropriate for this database. Mainly, because determining the underlying emotion would be much more challenging for dysarthric speech than it would be for typical speech, thereby adding more ambiguity to this unexplored domain. Also, natural recordings mostly suffer from poor quality, are often protected by copyright laws and privacy policies, and in addition suffer from lack of control on different aspects that are related to the recording settings which makes the task of processing the data very challenging (Scherer, 2003; (Busso et al., 2008). The acted methodology was also not appropriate as recruiting actors who had dysarthric speech was deemed to be difficult.

Table 1
Speakers' details (The dysarthria severity levels indicated in the table are informal judgments by the authors).

Speakers with dysarthria Type of dysarthria	Speaker	Gender	Age	Dysarthria severity/ time diagnosed
Spastic dysarthria (cerebral palsy)	DS01F	Female	65 years	Severe/from birth
Hypokinetic dysarthria (Parkinson's disease)	DS02F	Female	71 years	Mild/10 years
	DS03M	Male	66 years	Moderate/ 9 years
	DS04F	Female	68 years	Mild-to-moderate/ 10 years

Speakers with typical speech Gender	Number of speakers	Age Mean	SD	Range
All female	9	34.00	13.26	20–56
All male	12	35.67	16.81	19–70
Close in age female	1	56.00	–	56
Close in age male	2	66.00	5.66	62–70

This left us with the elicited approach. In order to elicit specific emotions in speakers, emotion stimuli has been chosen as the eliciting technique. Very short video clips of emotion stimuli were presented in order to elicit specific emotions. The video clips that have been used are adopted from those used when recording the Surrey Audio-Visual Expressed Emotion (SAVEE) database (Jackson and Haq, 2011). These video clips were taken from popular movies and television series. In addition to that, speakers were told that they can use Stanislavski's emotional memory techniques where they can remember the details of a situation with the same emotion if they think it will help them to put themselves in a particular emotional state (Stanislavsky et al., 1936). This follows standard protocols for recording such databases (Burkhardt et al., 2005; Jackson and Haq, 2011; Livingstone and Russo, 2018). Speakers were given time to put themselves into a specific emotional state. They were told that they could repeat a sentence as many times as they wanted until they felt satisfied with their performance. Speakers were explicitly instructed to provide genuine expressions of emotions as they would do in typical everyday scenarios. No instructions or guidance were given as to how a particular emotion should be expressed.

2.1.1. *Participants*

There were three groups of participants in this study: speakers with dysarthria associated with cerebral palsy, speakers with dysarthria associated with Parkinson's disease, and speakers with typical speech. All participants were recruited using advertising emails sent to special email lists, flyers, and word of mouth in the area of Sheffield, UK. The inclusion criteria for all of the three groups were that the participant must be a native British English speaker, over the age of 18, and have no known cognitive problems and no known literacy difficulties. None of the participants were professional actors. Informed consent was obtained from all participants.

Speakers with dysarthria. Two groups of participants with dysarthria were included in this study. The first group contained 1 female speaker with severe dysarthria associated with cerebral palsy. The second group contained 2 female speakers and 1 male speaker with dysarthria associated with Parkinson's disease. Recordings of speakers with Parkinson's disease were taken while they were under the anti-Parkinsonian medications effect. Table 1 lists the details of the speakers and their dysarthria severity levels.

Speakers with typical speech. Twenty-one speakers with typical speech were included in this study, 12 male and 9 female. Table 1 lists the details of the speakers of this group. Table 1 also lists individual details of speakers who are close in age to the speakers with dysarthria.

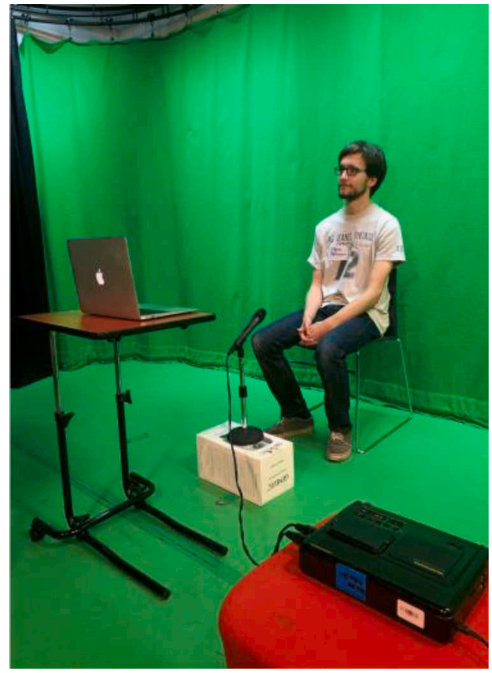


Fig.1. Data capture physical setup.

2.1.2. *Materials*

The recordings took place in a professional recording studio at the University of Sheffield. Fig. 1 shows the data capture physical setup. All speakers sat during the recording facing the camera. A green screen cloth was used as the background.

The microphone was placed approximately 50 cm from the speaker's mouth. A break of 5 to 10 min was given after finishing one complete set of the sentence blocks. All speakers completed their recordings in one session.

Since this analysis relies solely on speech, only the details of capturing the audio will be provided here. Speakers were recorded individually using Marantz PMD 670 recorder with the following settings: the recording level was set to 4, the sampling rate was set to 16,000 Hz, and mono audio channel was used in all the recording sessions. Audio files were saved in WAV format. The prompting material was displayed on a 13 inch Macbook Air placed on a Table 1 meter from the speaker. The audio file for each speaker was exported from the recorder and sentence segmentation was performed manually using Audacity software (http://audacity.sourceforge.net).

2.1.3. *Selection of emotions*

With respect to the set of emotions recorded, the widely adopted approach is to capture a small set of 'basic' emotions (Douglas-Cowie et al., 2003). Most of the discrete emotion models are taken from Darwin's "The Expression of Emotion in Man and Animals" (Darwin, 1872). This discrete emotion pattern approach has been popularized by scholars in this field: Tomkins, Ekman and Izard (Tomkins, 1962, 1963; Ekman, 1972; Van Bezooijen et al., 1983, 1980, 1992; Ekman et al., 1987; Izard, 1994, 1971; Levenson et al., 1992). In this study, a subset of the basic emotions has been included, namely, 'angry', 'happy', and 'sad'. 'Neutral' state has also been included as a baseline condition. The selection of these emotions was guided by several points:

- Given that this is a first of its kind study, starting with a smaller non-overlapping set can provide the base for a more focused initial exploration of the problem. In particular, this can allow us to answer the main question of whether or not people with dysarthria can convey emotions in their speech.

- Based on a survey the authors have conducted in order to understand emotion communication by people with dysarthria, ‘anger’, ‘happiness’ and ‘sadness’ were chosen by people with dysarthria as the most important emotions in terms of being able to communicate them successfully (Alhinti et al., 2020).
- This set of emotions are widely adopted in the literature when performing acoustic analysis (Yildirim et al., 2004; Kumbhakarn and Sathe-Pathak, 2015; Davletcharova et al., 2015).
- ‘Neutral’ was included in this study to be able to compare how different emotions affect speech compared to the neutral state.

2.1.4. Stimulus sets

The set of sentences is a subset of the sentences used in the SAVEE database (Jackson and Haq, 2011). Long sentences were excluded from the adopted set of sentences, as it might have been difficult for some people with dysarthria to be able to speak them. Although, a subset of the basic emotions has been included in this study, the original recordings of the data consisted of the full set of basic emotions, ‘angry’, ‘happy’, ‘sad’, ‘surprise’, ‘fear’, and ‘disgust’, in addition to the ‘neutral’ state. This gives a total of 7 emotional states. The full set consisted of 10 TIMIT sentences per emotion: 3 common, 2 emotion-specific and 5 generic sentences that were different for each emotion. The 3 common, and the $2 \times 6 = 12$ emotion-specific sentences were recorded as ‘neutral’ in addition to 2 ‘neutral’ sentences and 3 generic sentences. This gives a total of 20 ‘neutral’ sentences. Therefore, a total of 80 utterances per speaker is recorded. Only the sentences that were recorded in the chosen subset of emotions, a total of 50 sentences, were included in the analysis. The 50 sentences that were used in the analysis is listed in Appendix A.

2.2. Procedure

The full set of sentences were divided into fourteen blocks, each block contained 5 sentences from the same emotion, except for the neutral state where the block contained 10 sentences. Each recorded set began with a neutral block followed by one block from each emotion. This division was applied to avoid bias caused by speakers’ fatigue, and to ensure that speakers who had dysarthria and could not record the whole set of sentences would be able to record one set which included a subset of the sentences that covers all of the emotions.

The stimuli presentation consisted of three main stages: task presentation, emotive video presentation, and sentence presentation. In the task presentation stage, the emotional state that the speaker should perform next was presented as text on a screen in front of the speaker for around 2 s. The emotive video presentation stage consisted of playing a short emotive video clip to help the speaker elicit the target emotional state. Finally, in the sentence presentation stage, each sentence within the current block of sentences was presented on the screen individually.

The data will be publically available for research purposes in the near future. For more information, please see (10.15131/shef.data.10605536)

2.3. Acoustic analyses

A total of 50 utterances per speaker were included in the analysis, with each emotion consisting of 10 utterances, except for the ‘neutral’ which has 20 utterances. The acoustic features investigated in this analysis are: RMS energy, F0, speech rate, jitter, shimmer, and harmonic to noise ratio (HNR). All the acoustic features were extracted using the Praat tool (Boersma and Weenink, 2019), except for the RMS energy, which was extracted using Librosa, a python package for music and audio analysis (McFee et al., 2015). Default settings were chosen for all parameters unless specified otherwise. The choice of these features was guided by several points: 1) these features are among the most important and relevant features that show correlations with different vocal emotions expressions (Laukka et al., 2005; Kumbhakarn and Sathe-Pathak, 2015; Yildirim et al., 2004; Toivanen et al., 2006; Schuller et al., 2005; Kim et al., 2013), 2) all or part of these features are widely adopted in the literature with success for tasks related to analyzing the acoustic characteristics of emotional speech (Kumbhakarn and Sathe-Pathak, 2015; Yildirim et al., 2004; Toivanen et al., 2006; Schuller et al., 2005; Kim et al., 2013), and 3) all or part of these features have been included in some standardized sets developed for related tasks (Eyben et al., 2016; Schuller et al., 2009, 2013). Given that the purpose of this analysis is to mainly see whether the groups under study have enough control to communicate emotions through their voices or not, and to see how different their way is, compared to the typical speech control group, it is sufficient to start with a minimal set of potential acoustic features.

2.3.1. RMS energy

The root mean square energy is a common way to calculate the energy in a speech signal. It is calculated as the square root of the average sum of the squares of the amplitude of the signal samples. Research shows that high energy is usually associated with high-arousal emotions such as ‘angry’ and ‘happy’, while low energy is more associated with low-arousal emotions such as ‘sad’ (Johnstone and Scherer, 2000). The RMS energy of each utterance was computed using the utterance spectrogram with the following settings: 25 ms frame size and 10 ms overlap.

2.3.2. Fundamental frequency

Pitch is one of the most important perceptual features of sound that mainly depends on a sound’s frequency and F0 (Plack et al., 2014). The analysis of the F0 including minimum, maximum, range, mean, has been included on most of the research on emotional speech. F0 statistics are one of the most important features that correlate with emotional vocal expressions. Higher and wider range of F0 is usually associated with high-arousal emotions such as ‘angry’ and ‘happy’ compared to neutral speech while lower and less wider range of F0 is more associated with low-arousal emotions such as ‘sad’ (Guo et al., 2016; Murray and Arnott, 1993; Breitenstein et al., 2001; Johnstone and Scherer, 2000; Guo et al., 2016). In this study, the F0 contour and related F0 statistics were computed using the autocorrelation method through the *To Pitch* command in Praat with the following pitch range settings: from 60 to 500 Hz. The two statistics that have been analyzed under this feature are the F0 mean and range. For each utterance, the range of F0 was calculated by subtracting the minimum F0 from the maximum F0 values.

2.3.3. Speech rate

Speech rate is determined by the number of syllables spoken per time unit. It is an important feature that has been used in different tasks such as determining fluency in second language learning and determining the speaker’s emotional states. Research shows that speech rate has correlation with vocal arousal (Juslin et al., 2005). The experiment reported by Breitenstein et al. (2001) shows that slow speech rate is associated with ‘sad’ emotion while fast speech rate is associated with ‘angry’ and ‘happy’.

Speech rate per utterance was calculated using a Praat script where the syllable boundaries are estimated using energy-based syllable-nuclei detection method (De Jong and Wempe, 2009).

2.3.4. Jitter

In periodic signals, jitter shows how the signal deviates from its true periodicity. It is a measure of the fundamental frequency variations from cycle to cycle. There are several types of jitter measurements. In this analysis, the jitter local absolute (known as jitta) was chosen. It is the average absolute difference between consecutive periods represented in seconds and was computed using the *Get jitter (local, absolute)* command in Praat.

2.3.5. Shimmer

In periodic signals, shimmer shows the cycle to cycle variations of amplitude. There are also several types of shimmer measurements. In this analysis the shimmer local (dB) was chosen. It represents the

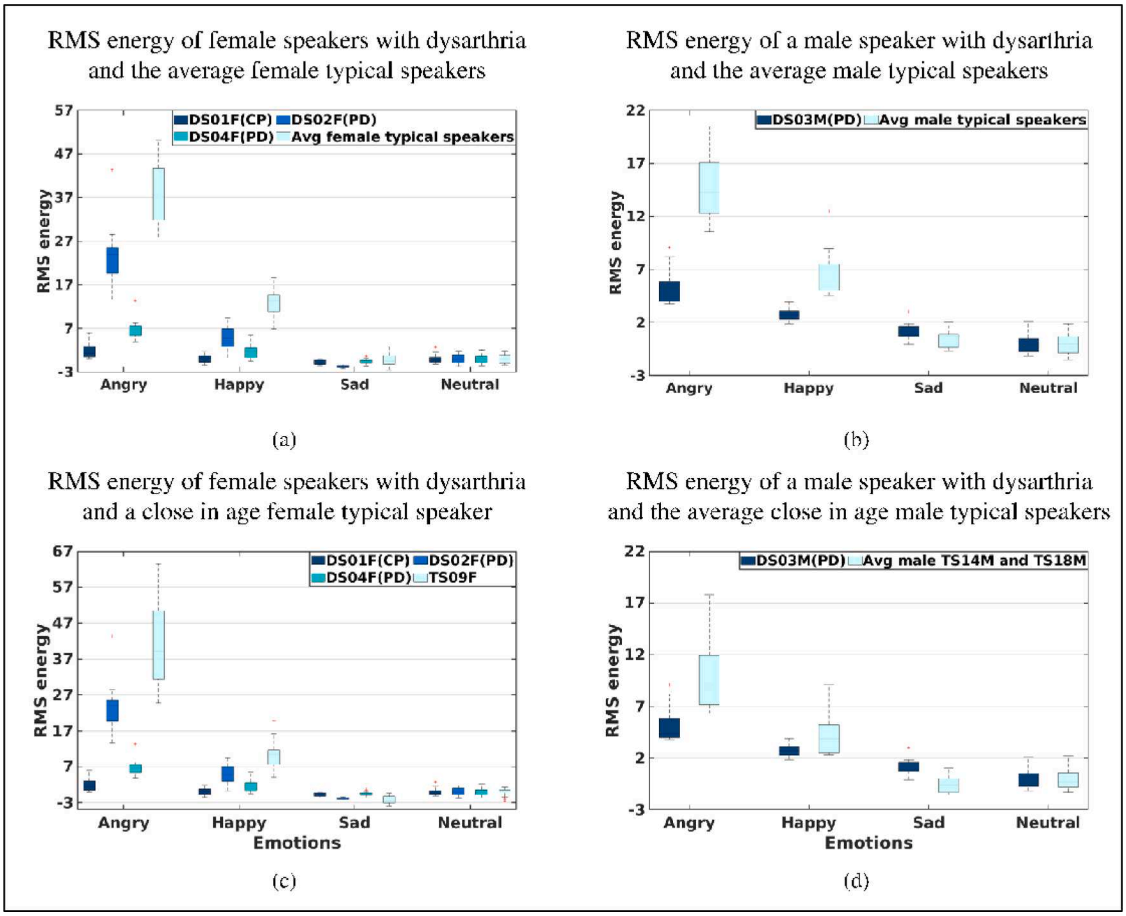


Fig. 2. Boxplot of the RMS energy of female and male speakers.

difference in peak to peak amplitude in decibels. It was computed using the *Get Shimmer (local_dB)* command in Praat.

Research has shown that jitter and shimmer are important features in emotion classification (Li et al., 2007; Hossain and Naznin, 2018; Juslin et al., 2005). Whiteside (1998) found that high jitter and shimmer are associated with high-arousal emotions, such as ‘angry’, while low levels are associated with low-arousal emotions such as ‘sad’.

2.3.6. Harmonics-to-noise ratio (HNR)

Harmonics-to-noise ratio is a measure of the additive noise in the voice signal. It is a useful feature to measure the breathiness and roughness (hoarseness) of a voice (krom, 1995). Research shows that HNR has higher values in negative emotions such as ‘anger’ compared to the ‘neutral’ state (Alter et al., 1999). HNR values were computed using Praat *To Harmonicity (cc)* command with minimum pitch set to 60 Hz.

3. Results

Acoustic analysis was performed on all 200 utterances (50 utterances x 4 speakers) produced by speakers with dysarthria and on all 1050 utterances (50 utterances x 21 speakers) produced by speakers with typical speech. For each feature, a linear multi-level model was used to analyze the data. The feature being analysed was the response variable. The fixed factors were the type of speech, (hereinafter referred to as condition; typical speech (TS), dysarthric speech associated with cerebral palsy (CP), dysarthric speech associated with Parkinson’s disease (PD)), gender (female, male), and emotion (‘angry’, ‘happy’, ‘sad’, ‘neutral’). The interaction between these fixed factors was computed, with speaker identity and sentence as random factors. Using estimated marginal means, pairwise comparison for the main effects and their interaction were conducted on each feature where the p values were adjusted using the Bonferroni method. Since F0 and RMS energy are known to differ between male and female speakers, the analyses of the related features were done separately in both plots and statistical models where gender was added in the interaction terms. As normal aging can affect some acoustic characteristics of speakers, we also compared speakers with dysarthria to closely age-matched speakers with typical speech by plotting the results.

3.1. RMS energy

Fig. 2 shows the boxplot of the RMS energy for female and male speakers after standardization. The RMS energy was standardized using the average energy of the ‘neutral’ state of each speaker/ group of speakers. The standardization was done to remove any effect of recording differences that can occur such as possible distance differences between speakers and the microphone. Fig. 2a and b, show that female speakers with dysarthria have lower RMS energy compared to the average for female typical speakers in all of the three emotions. This is also the case for the male speaker with dysarthria in ‘angry’ and ‘happy’ emotions. Although there is a difference in the range of energy produced by speakers with dysarthria compared to typical speakers, all speakers seemed to have the ability to vary energy when trying to communicate different emotional states. Table 2 illustrates the pairwise comparison for the main effects of condition (TS, CP, PD), gender (female, male), and emotions (‘angry’, ‘happy’, ‘sad’, ‘neutral’) on RMS energy corrected using Bonferroni adjustment. The table indicates that the main effect of condition reflects a significant difference ($p < 0.001$) between TS and CP and a significant difference ($p < 0.01$) between TS and PD, while the difference between CP and PD is not significant. In addition, the

Table 2
Pairwise comparison of the estimated marginal means on RMS energy of the main effects and interaction effect of (gender*condition*emotion) using multilevel modeling. F/Female, M/Male, A/Anger, H/Happy, S/Sad, and N/ Neutral. (Where * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$).

Fixed Factor	(i)		(j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
Condition	TS		CP	6.045	1.335	1250.000	***	2.845	9.245
			PD	3.014	0. 848	1250.000	**	0.981	5.048
	CP		PD	−3.030	1.529	1250.000		−6.694	0.634
Gender	F		M	1.507	0.874	1250.000		−0.208	3.222
Emotions	N		A	−11.648	1.121	1250.000	***	−14.609	−8.686
			H	−4.041	1.121	1250.000	**	−7.002	−1.079
			S	−0.185	1.121	1250.000		−3.146	2.776
	A		H	7.607	1.294	1250.000	***	4.187	11.026
			S	11.648	1.294	1250.000	***	8.043	14.882
	H		S	3.856	0.808	1250.000	*	0.436	7.275
Interaction effects				Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference	
Gender	Condition	Emotion (i)	Emotion (j)					Lower Bound	Upper Bound
F	TS	N	A	−25.682	1.138	1250.000	***	−28.689	−22.676
			H	−8.928	1.138	1250.000	***	−11.935	−5.922
			S	−1.112	1.138	1250.000		−4.118	1.895
		A	H	16.754	1.314	1250.000	***	13.282	20.226
			S	24.571	1.314	1250.000	***	21.009	28.042
		H	S	7.817	1.314	1250.000	***	4.345	11.289
	CP	N	A	−1.966	3.414	1250.000		−10.986	7.054
			H	−0.184	3.414	1250.000		−9.204	8.836
			S	0.796	3.414	1250.000		−8.224	9.816
		A	H	1.782	3.942	1250.000		−8.634	12.197
			S	2.762	3.942	1250.000		−7.654	13.178
		H	S	0.980	3.942	1250.000		−9.435	11.396
	PD	N	A	−15.631	2.414	1250.000	***	−22.009	−9.253
			H	−3.306	2.414	1250.000		−9.684	3.073
			S	1.159	2.414	1250.000		−5.219	7.538
		A	H	12.325	2.787	1250.000	***	4.960	19.690
			S	16.790	2.787	1250.000	***	9.425	24.155
		H	S	4.465	2.787	1250.000		−2.900	11.830
M	TS	N	A	−9.671	0.985	1250.000	***	−12.275	−7.067
			H	−5.012	0.985	1250.000	***	−7.616	−2.408
			S	−0.659	0.985	1250.000		−3.263	1.945
		A	H	4.659	1.138	1250.000	***	1.652	7.666
			S	9.012	1.138	1250.000	***	6.006	12.019
		H	S	4.353	1.138	1250.000	**	1.347	7.360
	PD	N	A	−5.288	3.414	1250.000		−14.308	3.732
			H	−2.773	3.414	1250.000		−11.793	6.247
			S	−1.110	3.414	1250.000		−10.130	7.910
		A	H	2.515	3.942	1250.000		−7.901	12.930
			S	4.178	3.942	1250.000		−6.238	14.594
		H	S	1.663	3.942	1250.000		−8.752	12.079

difference between females and males is not significant. The differences between all pairs of emotions are significant except between 'neutral'/ 'sad'. A significant difference ($p < 0.001$) between 'neutral'/'angry', 'angry'/'happy', and 'angry'/'sad', ($p < 0.01$) between 'neutral'/ 'happy', and a significant difference ($p < 0.05$) between 'happy'/'sad. 'Angry' has the highest mean estimates of RMS energy. Table 2 also illustrates the pairwise comparison for the interaction effect of gender, condition, and emotion on RMS energy corrected using Bonferroni adjustment. The Table shows a significant difference ($p < 0.001$) between all pairs of emotions except between 'neutral'/'sad' for both female and male typical speakers. In addition, a significant difference ($p < 0.001$) between 'neutral'/'angry', 'angry'/'happy', and 'angry'/ 'sad' for female speakers with PD. Although there are differences between the means in the other two groups (female speaker with CP and male speaker with PD), having only one speaker in each group means differences would have to be large to be considered significant. However, overall observations can still be made for these groups. 'Angry' has the highest mean estimates of RMS energy for all groups.

3.2. Fundamental frequency

Two statistics of F0 were analysed, the F0 range and mean F0. Fig. 3a, b, c, and d show the boxplot of F0 range and mean F0 per emotion for female and male speakers, respectively. Fig. 3e and f show the boxplot of mean F0 per emotion for female and male speakers, respectively, in comparison to close in age typical speakers. Fig. 4a and b show the F0 range of each utterance for the female speakers in the anger and neutral emotions, respectively.

Based on the conducted pairwise comparison for the main effects of condition, gender, and emotions on F0 range shown in Table 3, the following is observed: there is a significant difference ($p < 0.001$) between all conditions with higher mean estimates of F0 range for CP. The difference between females and males is significant ($p < 0.001$) with higher mean estimates for females. There is no significant difference between any pair of emotions. The interactions of gender, condition, and emotion on F0 range corrected using Bonferroni adjustment shown in Table 3, the difference between 'angry'/'sad' and 'happy'/'sad' is significant ($p < 0.01$) for female typical speakers. A significant difference ($p< 0.05$) between 'neutral'/'sad' and 'angry'/'sad for female speakers with PD. The difference between 'neutral'/'happy' and 'neutral'/'sad' is significant ($p < 0.01$) for male typical speakers. There is no significant difference detected between any pair of emotions for the female speaker with CP and the male speaker with PD.

The pairwise comparison for the main effects of condition, gender, and emotions on mean F0 shown in Table 4 indicates a significant difference ($p < 0.001$) between TS and CP and between CP and PD with higher mean estimates of mean F0 for CP. The difference between females and males is significant ($p < 0.001$) with females having the

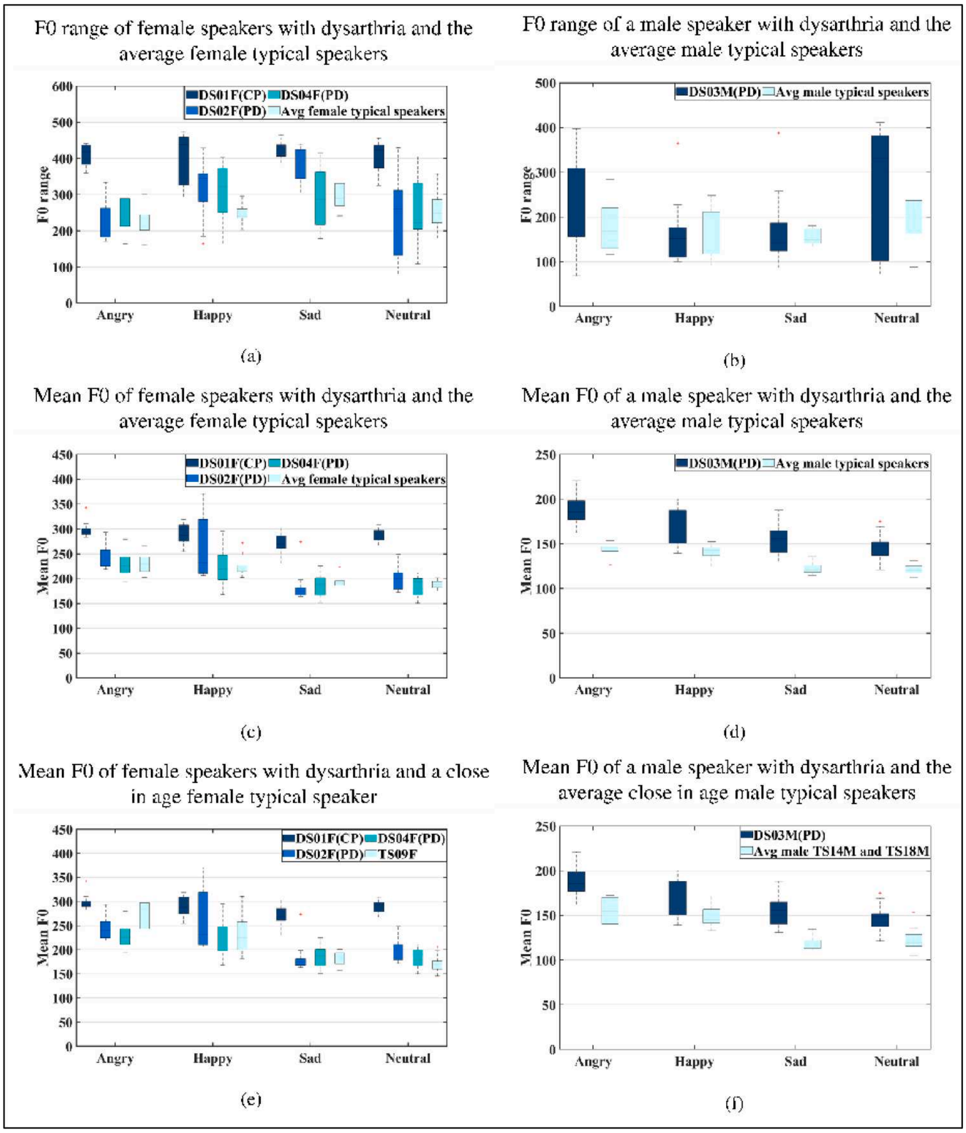


Fig. 3. F0 range and mean of female and male speakers with dysarthria and typical speakers.

highest marginal mean estimates. The differences between all pairs of emotions are significant ($p < 0.001$) except between ('neutral' and 'sad') and ('angry' and 'happy'). 'Angry' has the highest marginal mean estimates of F0 mean. Based on the conducted pairwise comparison for the interaction effect of gender, condition, and emotion on mean F0 presented in Table 4, the following is observed: there is a significant difference ($p < 0.001$) between all pairs of emotions except between 'neutral'/'sad' and 'angry'/'happy' for both female and male typical speakers and female speakers with PD. There is no significant difference detected between any pair of emotions for the female speaker with CP and the male speaker with PD except between 'neutral'/'angry' where a significant difference ($p<0.001$) is found for the male speaker with PD.

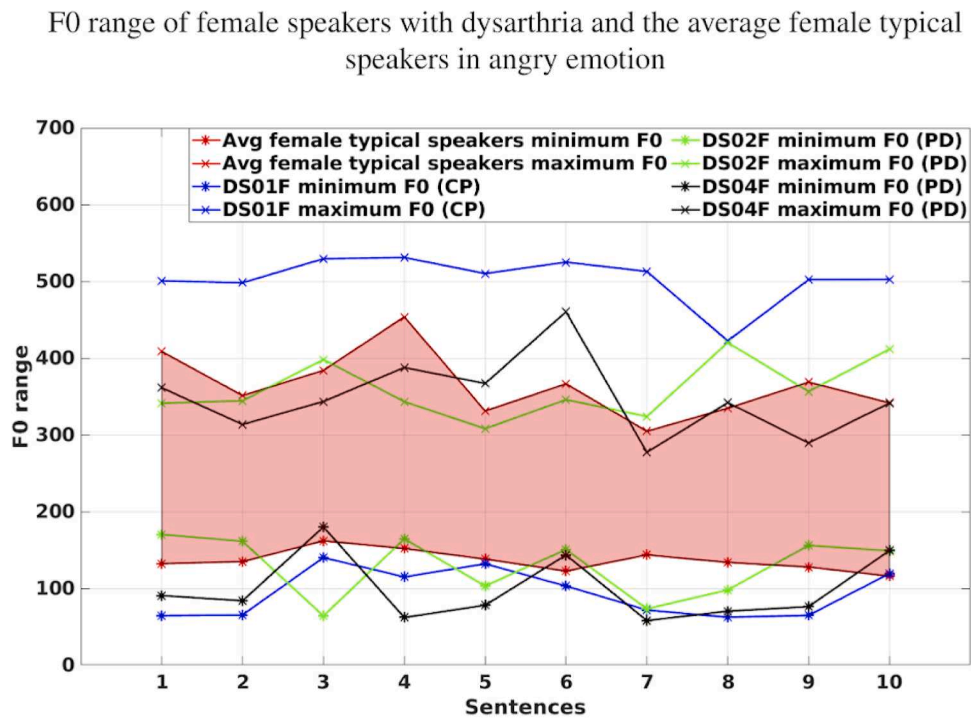
3.3. Speech rate

Fig. 5 shows the boxplot of the speech rate for each emotion. The speech rate for speaker DS01F who has severe dysarthria caused by cerebral palsy is much slower than other speakers with dysarthria caused by Parkinson's disease and typical speakers. This is actually expected due to the nature and severity level of speaker DS01F as she takes much longer time to articulate. The pairwise comparison for the main effects of condition, gender, and emotions on speech rate presented in Table 5, indicates a significant difference ($p < 0.001$) between TS and CP and between CP and PD with CP having the lowest marginal mean estimates. The difference between females and males is significant ($p < 0.01$) with higher marginal mean estimates for females. There is no significant difference between any pair of emotions. The pairwise comparison for the interaction effect of condition and emotions on speech rate presented in Table 5, indicates a significant difference ($p < 0.01$) between 'neutral'/'angry' and ($p < 0.001$) between 'neutral'/'sad' for the typical speakers group. There is no significant difference detected between any pair of emotions for the other groups.

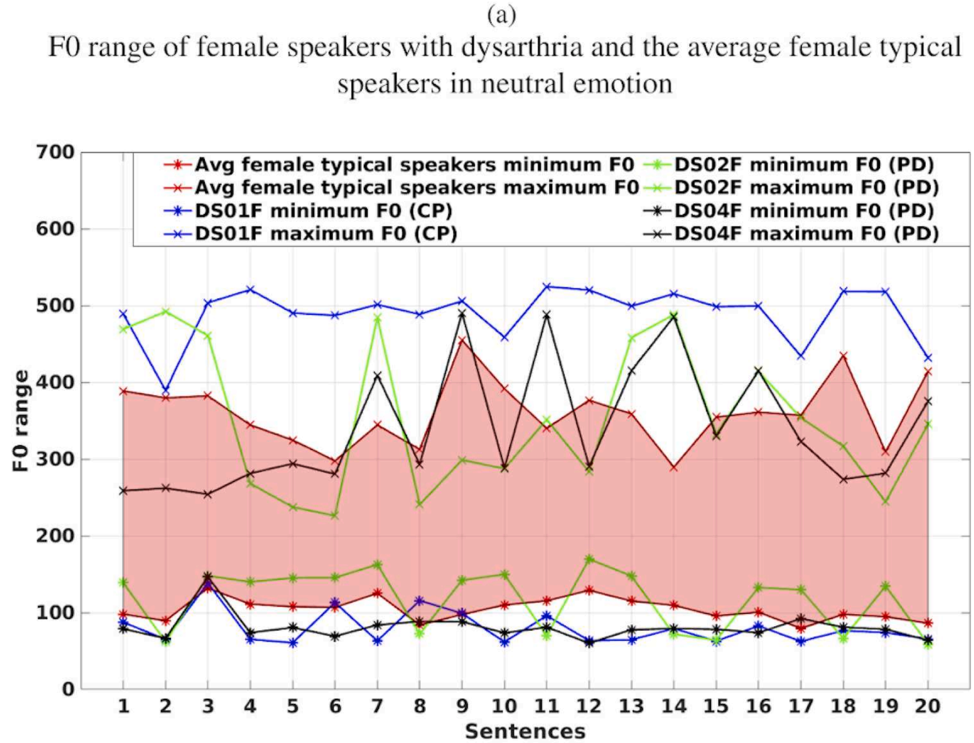
3.4. Jitter

Fig. 6a and b show the boxplot of the jitter local absolute feature of each emotion for female and male speakers, respectively. Fig. 6c and d show the values of the jitter local absolute feature of each emotion between speakers with dysarthria and close in age typical speakers.

The pairwise comparison for the main effects of condition, gender, and emotions on jitter local absolute presented in Table 6, indicates the significant main effect of condition reflects a significant difference ($p < 0.001$) between TS and CP and between CP and PD and ($p < 0.05$) between TS and PD. CP has the lowest marginal mean estimates. The



(a)



(b)

Fig. 4. F0 range of female speakers in (a) anger and (b) neutral emotions.

Table 3
Pairwise comparison of the estimated marginal means on F0 range of the main effects and interaction effect of (gender*condition*emotion) using multilevel modeling. F/Female, M/Male, A/Anger, H/Happy, S/Sad, and N/ Neutral. (Where * = p<0.05, ** = p<0.01, *** = p<0.001).

Fixed Factor	(i)	(j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
Condition	TS	CP	−227.535	19.036	1192.910	***	−273.170	−181.900
		PD	−64.824	13.783	849.565	***	−97.885	−31.762
	CP	PD	162.711	20.564	1248.621	***	113.415	212.007
Gender	F	M	134.722	11.803	1249.624	***	111.566	157.878
Emotions	N	A	22.622	15.989	1245.913		−19.630	64.873
		H	18.830	15.473	1249.972		−22.056	59.717
		S	−2.532	15.153	1248.843		−42.574	37.511
	A	H	−3.791	17.464	1248.579		−49.940	42.358
		S	−25.153	17.740	1249.938		−72.031	21.724
	H	S	−21.362	17.464	1248.579		−67.511	24.786

Interaction effects Gender	Condition	Emotion (i)	Emotion (j) (j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
F	TS	N	A	32.267	16.207	1246.284		−10.559	75.093
			H	12.344	15.697	1249.986		−29.136	53.823
			S	−37.587	15.383	1248.815		−78.236	3.061
		A	H	−19.923	17.729	1248.556		−66.773	26.927
			S	−69.854	18.001	1249.918	**	−117.422	−22.286
		H	S	−49.931	17.729	1248.556	**	−96.781	−3.081
	CP	N	A	4.155	46.144	1248.803		−117.778	126.088
			H	7.455	45.967	1248.236		−114.012	128.922
			S	−15.221	45.861	1247.837		−136.407	105.965
		A	H	3.300	52.944	1247.799		−136.605	143.204
			S	−19.376	53.036	1248.103		−159.523	120.770
		H	S	−22.676	52.944	1247.799		−162.580	117.228
	PD	N	A	11.008	32.853	1249.537		−75.804	97.819
			H	−62.628	32.604	1248.701		−148.783	32.528
			S	−90.019	32.454	1247.977	*	−175.777	−4.260
		A	H	−73.635	37.459	1247.903		−172.620	25.349
			S	−101.027	37.588	1248.471	*	−200.353	−1.700
		H	S	−27.391	37.459	1247.903		−126.375	71.593
M	TS	N	A	26.912	14.294	1241.557		−10.861	64.685
			H	48.121	13.714	1249.605	**	11.882	84.360
			S	44.689	13.353	1249.102	**	9.405	79.974
		A	H	21.209	15.381	1248.796		−19.434	61.851
			S	17.777	15.693	1249.989		−23.691	59.246
		H	S	−3.431	15.381	1248.796		−44.074	37.211
	PD	N	A	38.766	46.144	1248.803		−83.167	160.700
			H	88.861	45.967	1248.236		−32.606	210.328
			S	85.478	45.861	1247.837		−35.707	206.664
		A	H	50.094	52.944	1247.799		−89.810	189.999
			S	46.712	53.036	1248.103		−93.434	186.858
		H	S	−3.382	52.944	1247.799		−143.287	136.522

difference between females and males is significant ($p < 0.001$) with males having the highest marginal mean estimates. The differences between all pairs of emotions are significant ($p < 0.001$) except between (‘neutral’ and ‘sad’) and between (‘angry’ and ‘happy’). Based on the pairwise comparison for the interaction effect of gender, condition, and emotion on jitter local absolute shown in Table 6, the following is observed: the difference is significant ($p < 0.001$) between all pairs of emotions for the female typical speakers group except between ‘neutral’/‘sad’ and between ‘angry’/‘happy’. For the female speakers with PD, a significant difference ($p < 0.001$) is found between ‘angry’/ ‘sad’, ($p < 0.01$) between ‘neutral’/‘angry’ and ‘happy’/‘sad’, and (p<0.05) between ‘neutral’/‘sad’. The differences between all pairs of emotions for the male typical speakers are significant except between ‘angry’/‘happy’, with (p<0.001) for all the other pairs except between ‘happy’/‘sad’ where (p<0.01). A significant difference (p<0.01) between ‘neutral’/‘angry’ and (p<0.05) between ‘neutral’/‘happy’ is found for the male speaker with PD. There is no significant difference found between any pair of emotions for the female speaker with CP.

3.5. Shimmer

Fig. 7a and b show the boxplot of the shimmer local feature in dB of each emotion for female and male speakers, respectively. The pairwise comparison for the main effects of condition, gender, and emotions on shimmer local shown in Table 7 indicates no significant difference of the main effect condition. The difference between females and males is significant ($p < 0.05$) with males having the highest marginal mean estimates. The differences between all pairs of emotions are significant except between ‘neutral’/‘sad’ and between ‘angry’/‘happy’, where ($p <$ 0.001) for all the other pairs except between ‘happy’/‘sad’ where ($p <$ 0.01). From the pairwise comparison for the interaction effect of gender, condition, and emotion on shimmer local illustrated in Table 7 the following is observed: the differences between all pairs of emotions are significant (p< 0.001) for the female typical speakers except between ‘neutral’/‘sad’ and between ‘angry’/‘happy’. A significant difference (p<0.05) between ‘angry’/‘sad’ and between ‘happy’/‘sad’ for the female speakers with PD. For the male typical speakers, a significant difference (p<0.001) is found between ‘neutral’/ ‘angry’, ‘neutral’/ ‘happy’, and ‘neutral’/‘sad’. There is no significant difference found between any pair of emotions for the female speaker with CP and the male speaker with PD.

3.6. HNR

From Fig. 8, it can be seen that speaker DS01F has higher levels of HNR compared to the other speakers with dysarthria caused by

Table 4
Pairwise comparison of the estimated marginal means on mean F0 of the main effects and interaction effect of (gender*condition*emotion) using multilevel modeling. F/Female, M/Male, A/Anger, H/Happy, S/Sad, and N/ Neutral. (Where * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$).

Fixed Factor	(i)	(j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference	
							Lower Bound	Upper Bound
Condition	TS	CP	−106.251	4.290	1237.413	***	−84.940	−64.633
		PD	−6.645	3.120	1111.983		−14.124	0.835
	CP	PD	99.606	4.624	1248.349	***	88.521	110.961
Gender	F	M	86.550	2.655	1249.043	***	81.343	91.758
Emotions	N	A	−31.906	3.598	1249.867	***	−41.415	−22.398
		H	−25.688	3.481	1249.643	***	−34.885	−16.491
		S	−1.227	3.408	1248.455		−7.778	10.232
	A	H	6.218	3.927	1248.301		−4.159	16.595
		S	33.133	3.990	1249.337	***	22.589	43.677
	H	S	26.915	3.927	1248.301	***	16.538	37.293

Interaction effects				Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference	
Gender	Condition	Emotion (i)	Emotion (j) (j)					Lower Bound	Upper Bound
F	TS	N	A	−41.767	3.647	1249.900	***	−51.405	−32.130
			H	−36.942	3.531	1249.612	***	−46.273	−27.612
			S	−5.882	3.459	1248.438		−15.023	3.259
		A	H	4.825	3.987	1248.288		−5.710	15.360
			S	35.885	4.049	1249.306	***	25.186	46.584
		H	S	31.060	3.987	1248.288	***	20.525	41.595
	CP	N	A	−9.317	10.377	1248.430		−36.736	18.103
			H	1.295	10.336	1248.111		−26.018	28.609
			S	18.398	10.312	1247.902		−8.851	45.647
		A	H	10.612	11.905	1247.883		−20.846	42.070
			S	27.715	11.926	1248.040		−3.799	59.228
		H	S	17.103	11.905	1247.883		−14.355	48.560
	PD	N	A	−45.319	7.388	1248.932	***	−64.843	−25.796
			H	−48.718	7.332	1248.371	***	−68.092	−29.343
			S	6.790	7.297	1247.974		−12.493	26.074
		A	H	−3.398	8.423	1247.936		−25.655	18.859
			S	52.110	8.452	1248.240	***	29.775	74.445
		H	S	55.508	8.423	1247.936	***	33.251	77.765
M	TS	N	A	−21.038	3.218	1249.317	***	−29.540	−12.535
			H	−20.775	3.085	1249.871	***	−28.928	−12.622
			S	−1.490	3.003	1248.616		−9.425	6.444
		A	H	0.263	3.459	1248.427		−8.876	9.403
			S	19.547	3.530	1249.602	***	10.219	28.875
		H	S	19.248	3.459	1248.427	***	10.145	28.424
	PD	N	A	−42.089	10.377	1248.430	***	−69.509	−14.670
			H	−23.302	10.336	1248.111		−50.615	4.012
			S	−11.681	10.312	1247.902		−38.930	15.568
		A	H	18.788	11.905	1247.883		−12.670	50.246
			S	30.409	11.926	1248.040		−1.105	61.922
		H	S	11.621	11.905	1247.883		−19.837	43.079

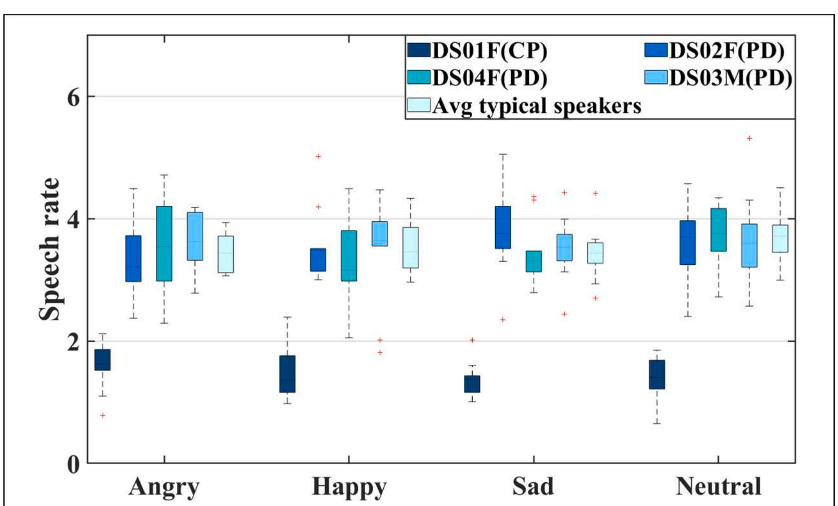


Fig. 5. Speech rate of speakers with dysarthria and the average typical speakers.

Table 5
Pairwise comparison of the estimated marginal means on speech rate of the main effects and interaction effect of (condition*emotion) using multilevel modeling. A/ Anger, H/Happy, S/Sad, and N/ Neutral. (Where * = p<0.05, ** = p<0.01, *** = p<0.001).

Fixed Factor	(i)	(j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
Condition	TS	CP	2.249	0.112	1082.441	***	1.979	2.518
		PD	0.149	0.076	357.635		−0.033	0.332
	CP	PD	−2.099	0.120	1249.811	***	−2.387	−1.812
Gender	F	M	0.132	0.042	1207.415	**	0.050	0.214
Emotions	N	A	0.048	0.110	1225.386		−0.242	0.339
		H	0.073	0.108	1247.902		−0.211	0.358
		S	0.107	0.106	1249.267		−0.173	0.387
	A	H	0.025	0.122	1248.858		−0.298	0.348
		S	0.058	0.124	1249.552		−0.268	0.385
	H	S	0.033	0.122	1248.858		−0.290	0.356

Interaction effects Gender	Condition	Emotion (i)	Emotion (j) (j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
All	TS	N	A	0.219	0.66	1024.132	**	0.043	0.395
			H	0.119	0.62	1194.727		−0.046	0.284
			S	0.242	0.60	1249.214	***	0.84	0.399
		A	H	−0.100	0.069	1249.949		−0.281	0.081
			S	0.022	0.071	1221.308		−0.081	0.210
		H	S	0.123	0.69	1249.949		−0.059	0.304
	CP	N	A	−0.206	0.271	1249.679		−0.922	0.511
			H	−0.075	0.270	1248.628		−0.789	0.639
			S	0.024	0.270	1247.632		−0.688	0.737
		A	H	0.131	0.311	1247.529		−0.692	0.953
			S	0.230	0.312	1248.316		−0.593	1.054
		H	S	0.100	0.311	1247.529		−0.723	0.922
	PD	N	A	0.132	0.159	1247.614		−0.287	0.551
			H	0.176	0.157	1249.959		−0.238	0.591
			S	0.054	0.156	1248.335		−0.358	0.466
		A	H	0.044	0.180	1248.071		−0.431	0.520
			S	−0.078	0.181	1249.670		−0.555	0.400
		H	S	−0.122	0.180	1248.071		−0.597	0.353

Parkinson's disease and the average typical speakers. From Table 8 we can see that based on the pairwise comparison for the main effects of condition, gender, and emotions on HNR, the significant main effect of condition reflects a significant difference ($p < 0.001$) between TS and CP and between CP and PD, but not between TS and PD. CP has the highest marginal mean estimates. The difference between females and males is significant ($p < 0.001$) with females having the highest marginal mean estimates. The differences between all pairs of emotions are not significant except between 'neutral'/'angry' and 'angry'/'sad' with ($p < 0.01$). From the pairwise comparison of the interaction effect of condition and emotion on HNR illustrated in Table 8, the following is observed: for typical speakers, the difference is significant ($p< 0.001$) between 'neutral'/'angry', 'neutral'/'happy', and 'neutral'/'sad' and ($p<0.01$) between 'angry'/'sad' but not between 'neutral'/'sad' and 'happy'/'sad'. There is no significant difference between any pair of emotions for the speaker with CP and the speakers with PD, except between 'angry'/'sad' with ($p<0.05$) for the former.

4. Discussion

The results show that some people with dysarthria, even severe dysarthria, are able to control some aspects of the suprasegmental and prosodic features of their speech to communicate emotions.

Although no strong conclusions can be made and significant difference between emotions and groups can be difficult to observe due to the very limited number of speakers in some of the groups (1 speaker in some of them), we can still make observations. The changes to these features appear similar to those of typical speakers, despite speakers with dysarthria having a more limited articulatory and prosodic control. It is likely that these systematic changes help in the communication of the emotions.

One of the features used to distinguish emotion is the RMS energy. Typical speakers vary it while expressing the emotions investigated in this study, except between 'neutral'/'sad' utterances. Female speakers with PD also managed to vary it significantly when expressing some of the emotions. Similar to typical speakers, all speakers with dysarthria produced higher RMS energy when communicating high-arousal emotions such as 'angry' and 'happy' compared to low-arousal emotions such as 'sad' and 'neutral' [as can be seen from Fig. 2 and from the statistical model presented in Table 2]. Despite that, the differences between some high-arousal emotions and low-arousal emotions were not marked as significant for some of the groups due to the limited number of observation. However, the differences between the marginal means are still observed and with more data, significancy might be confirmed. This aligns with the findings reported in the literature on typical speech (Johnstone and Scherer, 2000). There is a significant difference in the RMS energy between the group of speakers with dysarthria caused by cerebral palsy and the typical speakers, and between the group of speakers with dysarthria caused by PD and the group of typical speakers.

The range of F0 does not appear to be a strong distinguishing feature. A significant difference was found between the three groups of speakers in the range of F0.

The mean F0 is an important feature in distinguishing emotions. The mean F0 can be used to distinguish between high-arousal and low-arousal emotions as the difference differs significantly [as can be seen from Table 4]. This significant difference was also observed within each group with more than one speaker from the results of the interaction effect. This is also consistent with the findings in the literature for speakers with typical speech, where high F0 is usually associated with high-arousal emotions and low F0 is more associated with low-arousal emotions (Breitenstein et al., 2001; Johnstone and Scherer, 2000; Guo et al., 2016). From the effect of condition, a significant difference was found in the mean F0 between the speaker with dysarthria caused by cerebral palsy and the other two groups of speakers.

From our analysis, it is observed that the speech rate is not vary

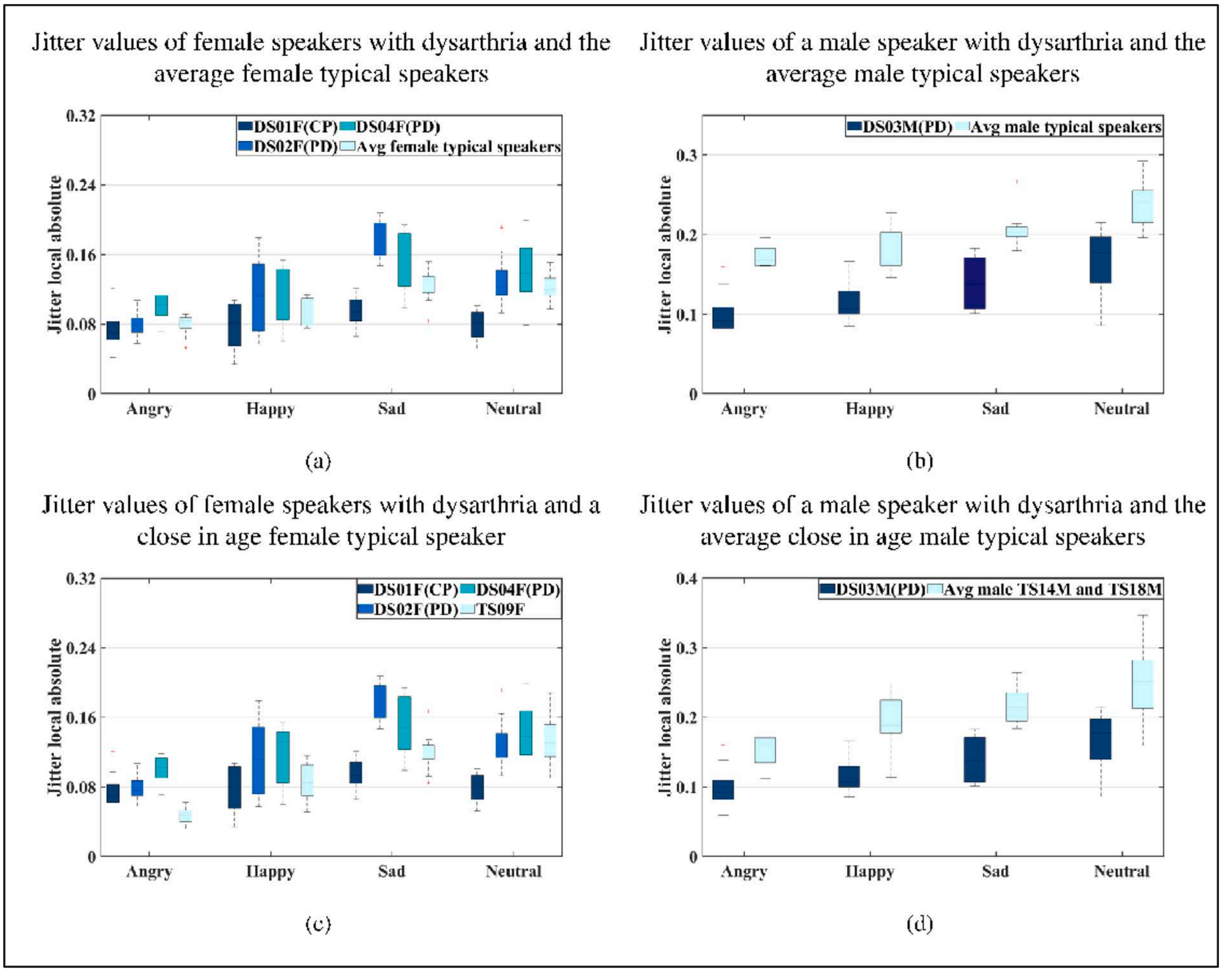


Fig. 6. Jitter values of (a) female and (b) male speakers.

useful in distinguishing between pairs of emotions as the differences between pairs of emotion were not statistically significant except between 'neutral'/'angry' and 'neutral'/'sad' for typical speakers. A significant difference was also found in the speech rate between the speaker with dysarthria caused by cerebral palsy and the other two groups of speakers.

The jitter local absolute is also a useful feature in distinguishing pairs of emotions. It can be used to distinguish between high-arousal and low-arousal emotions. The differences in the marginal means between these pairs of emotions were marked significantly within each group with more than one speaker. It is also observed that differences in the mean values of the jitter between the three groups of speakers: typical speakers, speakers with dysarthria caused by cerebral palsy and speakers with dysarthria caused by PD, were statistically significant.

Shimmer can be used to distinguish some pairs of emotions. As can be seen from the results of the interaction effect shown in Table 7, these pairs vary among groups. For example, while shimmer can be used for both female typical speakers and female speakers with dysarthria caused by PD to differentiate between 'anger'/'sad' and between 'happy'/'sad', where the mean difference between these emotions were found to be significant, in addition to other pairs of emotions for the group of female typical speakers, it can be used to distinguish neutral from all the other three emotions for the group of male typical speakers. There is no significant difference found for the effect of condition.

For typical speakers, the HNR feature appear to be sufficient to distinguish between pairs of emotions in our data except between 'angry'/'happy' and 'happy'/'sad'. No other significant difference was found in the HNR for the other groups except between 'angry'/'sad' for the speaker with severe dysarthria caused by cerebral palsy. From the effect of condition, a significant difference was found in the HNR between the speaker with dysarthria caused by cerebral palsy and the other two groups of speakers.

In our analysis, speaker DS01F, who has severe dysarthria due to cerebral palsy, has either higher values such as in F0 range and HNR or lower values such as in speech rate than the other female speakers with dysarthria caused by Parkinson's disease and the average female speakers with typical speech. This difference may be due to speaker DS01F having severe dysarthria, in contrast to the other speakers who have mild dysarthria. It is also observed that the characteristics of speakers with dysarthria caused by Parkinson's disease differ in some of the cases from those with typical speech. In addition, there is inter-speaker variation observed between the speakers with dysarthria caused by Parkinson's disease. This inter-speaker variation complies with the findings reported in the literature (Ma et al., 2010; Liu et al., 2019).

As the aim of this study is to know whether speakers with dysarthria have the ability to control some acoustic features while communicating emotions, some potential features have been analysed and the effect of a number of factors (condition, gender, and emotions) and their interaction on each feature has been investigated.

Classifying emotions from speech is by itself a challenging problem (El Ayadi et al., 2011). In the case of having disordered speech, this may be a more difficult classification problem as the speakers often have less control of the signifying features. Our analysis demonstrates the existence of significant differences between emotions in some of the investigated features. Yet, it is still unclear whether this level of difference is

Table 6
Pairwise comparison of the estimated marginal means on jitter local absolute of the main effects and interaction effect of (gender*condition*emotion) using multilevel modeling. F/Female, M/Male, A/Anger, H/Happy, S/Sad, and N/ Neutral. (Where * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$).

Fixed Factor	(i)	(j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
Condition	TS	CP	0.065	0.008	729.724	***	0.046	0.084
		PD	0.016	0.006	165.417	*	0.002	0.029
	CP	PD	−0.049	0.009	1249.518	***	−0.070	−0.029
Gender	F	M	−0.059	0.005	1248.585	***	−0.069	−0.050
Emotions	N	A	0.044	0.007	1081.311	***	0.026	0.061
		H	0.030	0.006	1218.991	***	0.013	0.047
		S	0.000	0.006	1249.988		−0.016	0.017
	A	H	−0.014	0.007	1249.839		−0.033	0.006
		S	−0.043	0.007	1235.860	***	−0.061	−0.024
	H	S	−0.030	0.007	1249.839	***	−0.049	−0.011

Interaction effects Gender	Condition	Emotion (i)	Emotion (j) (j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
F	TS	N	A	0.040	0.007	1090.147	***	0.023	0.058
			H	0.028	0.006	1221.245	***	0.011	0.045
			S	−0.002	0.006	1249.998		−0.018	0.015
		A	H	−0.012	0.007	1249.802		−0.032	0.007
			S	−0.042	0.007	1237.039	***	−0.061	−0.022
		H	S	−0.030	0.007	1249.802	***	−0.049	−0.010
	CP	N	A	0.004	0.019	1250.000		−0.047	0.054
			H	0.002	0.019	1248.988		−0.048	0.053
			S	−0.015	0.019	1247.326		−0.065	0.036
		A	H	−0.001	0.022	1247.138		−0.059	0.056
			S	−0.018	0.022	1248.506		−0.076	0.040
		H	S	−0.017	0.022	1247.138		−0.075	0.041
	PD	N	A	0.044	0.014	1246.532	**	0.008	0.080
			H	0.019	0.013	1249.971		−0.017	0.055
			S	−0.037	0.013	1247.977	*	−0.072	−0.001
		A	H	−0.025	0.015	1247.643		−0.066	0.016
			S	−0.081	0.016	1249.640	***	−0.122	−0.040
		H	S	−0.056	0.015	1247.643	**	−0.097	−0.015
M	TS	N	A	0.065	0.006	992.912	***	0.050	0.081
			H	0.051	0.006	1193.446	***	0.036	0.066
			S	0.029	0.006	1249.538	***	0.014	0.044
		A	H	−0.014	0.006	1250.000		−0.031	0.003
			S	−0.036	0.006	1221.968	***	−0.053	−0.019
		H	S	−0.022	0.006	1250.000	**	−0.039	−0.005
	PD	N	A	0.065	0.019	1250.000	**	0.015	0.116
			H	0.051	0.019	1248.988	*	0.001	0.101
			S	0.026	0.019	1247.326		−0.025	0.076
		A	H	−0.015	0.022	1247.138		−0.072	0.043
			S	−0.040	0.022	1248.506		−0.098	0.018
		H	S	−0.025	0.022	1247.138		−0.083	0.033

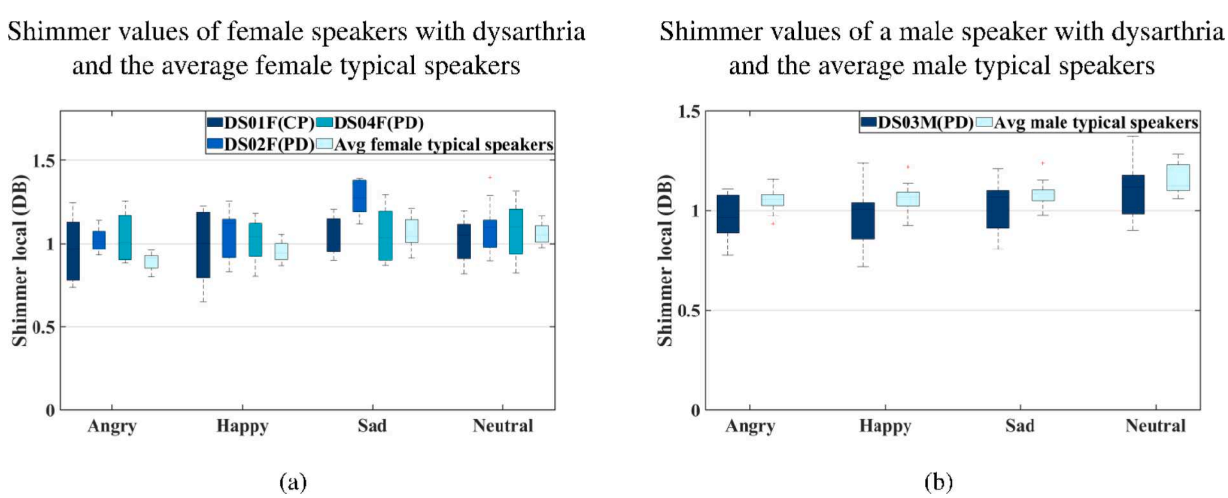


Fig. 7. Shimmer values in DB of (a) female and (b) male speakers.

Table 7
Pairwise comparison of the estimated marginal means on shimmer local of the main effects and interaction effect of (gender*condition*emotion) using multilevel modeling. F/Female, M/Male, A/Anger, H/Happy, S/Sad, and N/ Neutral. (Where * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$).

Fixed Factor	(i)	(j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
Condition	TS	CP	0.039	0.022	1250.000		−0.014	0.092
		PD	−0.011	0.014	1250.000		−0.044	0.023
	CP	PD	−0.050	0.025	1250.000		−0.110	0.011
Gender	F	M	−0.032	0.015	1250.000	*	−0.061	−0.003
Emotions	N	A	0.103	0.019	1250.000	***	0.054	0.152
		H	0.082	0.019	1250.000	***	0.033	0.132
		S	0.013	0.019	1250.000		−0.036	0.063
	A	H	−0.021	0.022	1250.000		−0.078	0.036
		S	−0.090	0.022	1250.000	***	−0.146	−0.033
	H	S	−0.069	0.022	1250.000	**	−0.126	−0.012

Interaction effects Gender	Condition	Emotion (i)	Emotion (j) (j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
F	TS	N	A	0.165	0.019	1250.000	***	0.115	0.215
			H	0.110	0.019	1250.000	***	0.060	0.160
			S	0.004	0.019	1250.000		−0.046	0.054
		A	H	−0.054	0.022	1250.000		−0.112	0.003
			S	−0.161	0.022	1250.000	***	−0.219	−0.103
		H	S	−0.106	0.022	1250.000	***	−0.164	−0.049
	CP	N	A	0.050	0.057	1250.000		−0.101	0.200
			H	0.034	0.057	1250.000		−0.116	0.184
			S	−0.001	0.057	1250.000		−0.151	0.149
		A	H	−0.015	0.066	1250.000		−0.189	0.158
			S	−0.015	0.066	1250.000		−0.224	0.123
		H	S	−0.035	0.066	1250.000		−0.209	0.138
	PD	N	A	0.057	0.040	1250.000		−0.049	0.163
			H	0.045	0.040	1250.000		−0.61	0.151
			S	−0.085	0.040	1250.000		−0.191	0.021
		A	H	−0.012	0.046	1250.000		−0.135	0.110
			S	−0.142	0.046	1250.000	*	−0.264	−0.019
		H	S	−0.129	0.046	1250.000	*	−0.252	−0.007
M	TS	N	A	0.104	0.016	1250.000	***	0.060	0.147
			H	0.097	0.016	1250.000	***	0.054	0.140
			S	0.070	0.016	1250.000	***	0.027	0.114
		A	H	−0.006	0.019	1250.000		−0.056	0.044
			S	−0.033	0.019	1250.000		−0.083	0.017
		H	S	−0.027	0.019	1250.000		−0.077	0.023
	PD	N	A	0.140	0.057	1250.000		−0.010	0.290
			H	0.125	0.057	1250.000		−0.072	0.275
			S	0.078	0.057	1250.000		−0.072	0.228
		A	H	−0.015	0.066	1250.000		−0.188	0.159
			S	−0.062	0.066	1250.000		−0.235	0.112
		H	S	−0.047	0.066	1250.000		−0.220	0.126

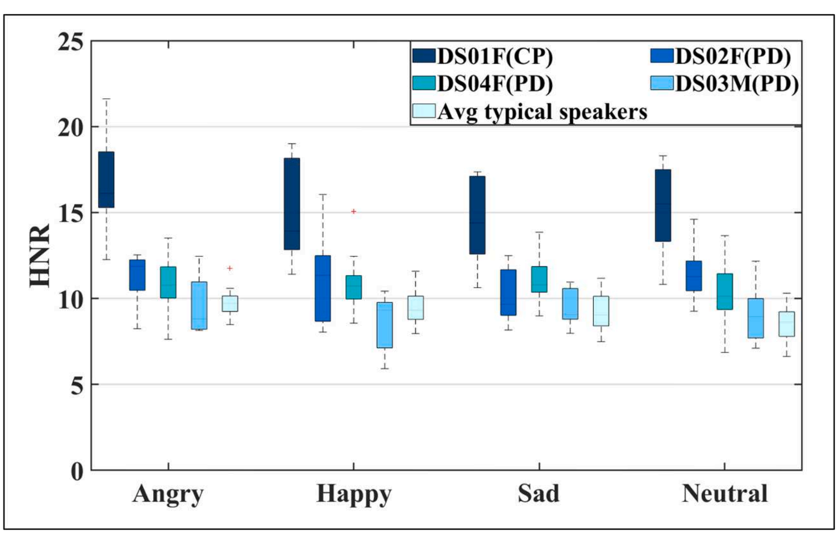


Fig. 8. HNR of speakers with dysarthria and the average typical speakers.

Table 8
Pairwise comparison of the estimated marginal means on HNR of the main effects and interaction effect of (condition*emotion) using multilevel modeling. A/Anger, H/Happy, S/Sad, and N/ Neutral. (Where * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$).

Fixed Factor	(i)	(j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
Condition	TS	CP	−4.941	0.294	1243.882	***	−5.647	−4.236
		PD	−0.069	0.203	1108.300		−0.555	0.417
	CP	PD	4.872	0.312	1248.657	***	4.123	5.621
Gender	F	M	0.882	0.109	1249.583	***	0.668	1.096
Emotions	N	A	−0.998	0.287	1249.996	**	−1.757	−0.239
		H	−0.267	0.280	1249.338		−1.008	−0.474
		S	−0.046	0.276	1248.303		−0.684	−0.776
	A	H	0.731	0.319	1248.183		−0.111	1.573
		S	1.044	0.322	1249.042	**	0.193	1.896
	H	S	0.313	0.319	1248.183		−0.528	1.155

Interaction effects Gender	Condition	Emotion (i)	Emotion (j) (j)	Mean Difference (i-j)	SE	df	p value	95% Confidence Interval for Difference Lower Bound	Upper Bound
All	TS	N	A	−1.275	0.174	1244.435	***	−1.736	−0.814
			H	−1.070	0.163	1249.764	***	−1.501	−0.639
			S	−0.646	0.156	1249.130	***	−1.057	−0.235
		A	H	0.205	0.179	1248.848		−0.268	0.678
			S	0.629	0.185	1249.999	**	0.139	1.118
		H	S	0.424	0.179	1248.848		−0.049	0.897
	CP	N	A	−1.469	0.707	1248.459		−3.336	0.398
			H	0.244	0.704	1248.123		−1.615	2.104
			S	0.776	0.702	1247.903		−1.079	2.631
		A	H	1.713	0.811	1247.882		−0.429	3.855
			S	2.245	0.812	1248.049	*	0.099	4.390
		H	S	0.532	0.811	1247.882		−1.610	2.674
	PD	N	A	−0.251	0.414	1249.371		−1.344	0.842
			H	0.024	0.409	1248.645		−1.056	1.104
			S	0.009	0.406	1248.053		−1.064	1.081
		A	H	0.275	0.469	1247.994		−0.963	1.513
			S	0.260	0.471	1248.455		−0.985	1.504
		H	S	−0.015	0.469	1247.994		−1.254	1.223

enough for people to accurately perceive these emotions. Assessing the importance of these features and how this relates to *perceived* emotions for listeners (familiar and unfamiliar with the participant) will also be the focus of our next study.

5. Conclusion

In this paper, we investigated whether people with dysarthria, caused by cerebral palsy and Parkinson's disease, are able to communicate emotions in their speech. We compared a set of acoustic features of the two types of dysarthria under this study to those of typical speech. We have also analysed the effect of different factors on each feature. Although the conducted analysis has the limitations of having been carried out on a limited number of speakers with dysarthria and using a limited number of sentences, it does, however, show that these people may have enough control to communicate intentions, gain attention, and convey emotions. This level of control of articulatory and prosodic features may not only help to train listeners to better recognize the emotions of speakers with dysarthria, but also to improve communication aids in a way that makes it more sensitive to specific cues in the vocalization signal produced by the speaker with dysarthria and act according to the speaker's intention. An example of a communication aid that can benefit from these findings is the voice-input voice-output (VIVOCA) project, which developed a communication aid that recognizes the disordered speech and reproduces it in a synthesized voice (Hawley et al., 2012).

The systematic changes made by people with dysarthria to communicate their emotional state are likely to help listeners understand the speaker emotion, and are also potentially useful for automatically classifying emotional state of a speaker.

CRediT authorship contribution statement

Lubna Alhinti: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft. **Heidi Christensen:** Supervision, Writing - review & editing, Methodology. **Stuart Cunningham:** Supervision, Writing - review & editing, Formal analysis, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been supported by a grant from the Saudi Ministry of Education, King Saud University, Saudi Arabia. The authors would like to thank the speakers whose participation in the recordings made this study possible. The authors would also like to thank the reviewers for their extremely helpful comments and efforts towards improving the manuscript.

Appendix A

Table A.1.

Table A.1
List of the sentences used in the recording and included in the analysis.

Emotion	Type of sentence	Sentence
Angry	Common	She had your dark suit in greasy wash water all year.
		Don't ask me to carry an oily rag like that.
		Will you tell me why?
	Emotion specific	Who authorized the unlimited expense account?
		Destroy every file related to my audits.
	Generic	Cory and Trish played tag with beach balls for hours.
		He will allow a rare lie.
		Withdraw all phony accusations at once.
		Right now may not be the best time for business mergers.
		A few years later the dome fell in.
Happy	Common	She had your dark suit in greasy wash water all year.
		Don't ask me to carry an oily rag like that.
		Will you tell me why?
	Emotion specific	Those musicians harmonize marvelously.
		The eastern coast is a place for pure pleasure and excitement.
	Generic	Project development was proceeding too slowly.
		The oasis was a mirage.
		Are your grades higher or lower than Nancy's?
		Serve the coleslaw after I add the oil.
		He would not carry a brief case.
Sad	Common	She had your dark suit in greasy wash water all year.
		Don't ask me to carry an oily rag like that.
		Will you tell me why?
	Emotion specific	The prospect of cutting back spending is an unpleasant one for any governor.
		The diagnosis was discouraging; however, he was not overly worried.
	Generic	Before Thursday's exam, review every formula.
		They enjoy it when I audition.
		John cleans shellfish for a living.
		He stole a dime from a beggar.
		American newspaper reviewers like to call his plays nihilistic.
Neutral	Common	She had your dark suit in greasy wash water all year.
		Don't ask me to carry an oily rag like that.
		Will you tell me why?
	Emotion specific	Who authorized the unlimited expense account?
		Destroy every file related to my audits.
		Please take this dirty table cloth to the cleaners for me.
		The small boy put the worm on the hook.
		Call an ambulance for medical assistance.
		Tornadoes often destroy acres of farm land.
		The carpet cleaners shampooed our oriental rug.
		His-shoulder felt as if it were broken.
		The prospect of cutting back spending is an unpleasant one for any governor.
		The diagnosis was discouraging; however, he was not overly worried.
		Those musicians harmonize marvelously.
		The eastern coast is a place for pure pleasure and excitement.
		The best way to learn is to solve extra problems.
		Calcium makes bones and teeth strong.
	Generic	Greg buys fresh milk each weekday morning.
		He always seemed to have money in his pocket.
		No return address whatsoever.

References

Alhinti, L., Christensen, H., Cunningham, S., 2020. An exploratory survey questionnaire to understand what emotions are important and difficult to communicate for people with dysarthria and their methodology of doing that. In: World Academy of Science, Engineering and Technology, International Conference on Bioinformatics Research and Applications.

Alter, K., Rank, E., Kotz, S.A., Pfeifer, E., Besson, M., Friederici, A.D., Matiasek, J., 1999. On the relations of semantic and acoustic properties of emotions. In: 14th International Congress of Phonetic Sciences. University of California, pp. 2121–2125.

Ansel, B.M., Kent, R.D., 1992. Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. J. Speech Lang. Hear. Res. 35 (2), 296–308.

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. 70 (3), 614.

Beukelman, D.R., Fager, S., Ball, L., Dietz, A., 2007. Aac for adults with acquired neurological conditions: a review. Augment. Alternative Commun. 23 (3), 230–242.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.07, retrieved 26 November 2019 from http://www.praat.org/.

Breitenstein, C., Lancker, D.V., Daum, I., 2001. The contribution of speech rate and pitch variation to the perception of vocal emotions in a german and an american sample. Cogn. Emot. 15 (1), 57–79.

Bunton, K., Weismer, G., 2001. The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. J. Speech Lang. Hear. Res.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., Weiss, B., 2005. A database of German emotional speech. In: Ninth European Conference on Speech Communication and Technology.

Busso, Carlos, Murtaza, Bulut, Chi-Chun, Lee, Abe, Kazemzadeh, Emily, Mower, Samuel, Kim, Jeannette N, Chang, Sungbok, Lee, Shrikanth S., Narayanan, 2008. IEMOCAP: Interactive emotional dyadic motion capture database. Language resources and evaluation 42 (4), 335. https://doi.org/10.1007/s10579-008-9076-6.

Calero, H.H., 2005. The Power of Nonverbal communication: How you Act is More Important Than What You Say. Silver Lake Publishing.

Canter, G.J., 1963. Speech characteristics of patients with Parkinson's disease: I. Intensity, pitch, and duration. J. Speech Hear. Disord. 28 (3), 221–229.

Connaghan, K.P., Patel, R., 2017. The impact of contrastive stress on vowel acoustics and intelligibility in dysarthria. J. Speech Lang. Hear. Res. 60 (1), 38–50.

Darley, F.L., Aronson, A.E., Brown, J.R., 1969. Differential diagnostic patterns of dysarthria. J. Speech Hear. Res. 12 (2), 246–269.

Darley, F.L., Aronson, A.E., Brown, J.R., 1975. Motor Speech Disorders. Saunders.

Darwin, C., 1872. The Expression of the Emotions in Man and Animal. The Portable Darwin, pp. 364–393.

Davletcharova, A., Sugathan, S., Abraham, B., James, A.P., 2015. Detection and analysis of emotion from speech signals. Procedia Comput. Sci. 58, 91–96.

De Jong, N.H., Wempe, T., 2009. Praat script to detect syllable nuclei and measure speech rate automatically. Behav. Res. Methods 41 (2), 385–390.

Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P., 2003. Emotional speech: towards a new generation of databases. Speech Commun. 40 (1–2), 33–60.

Duffy, J.R., 2013. Motor Speech disorders-e-book: Substrates, Differential diagnosis, and Management. Elsevier Health Sciences.

Ekman, P., 1972. Universal and cultural differences in facial expression of emotion. In: Nebraska Symposium on Motivation, 19, pp. 207–284.

Ekman, P., 1980. The face of man: expressions of universal emotions in a new guinea village. Scholarly Title.

Ekman, P., 1992. Facial Expressions of emotion: New findings, New Questions. SAGE Publications Sage CA, Los Angeles, CA.

Ekman, P., Friesen, W.V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., ..., 1987. Universals and cultural differences in the judgments of facial expressions of emotion. J. Pers. Soc. Psychol. 53 (4), 712.

El Ayadi, M., Kamel, M.S., Karray, F., 2011. Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recognit. 44 (3), 572–587.

Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., André, E., Busso, C., ..., 2016. The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. IEEE Trans. Affect. Comput. 7 (2), 190–202.

Ghio, A., Robert, D., Grigoli, C., Mas, M., Delooze, C., Mercier, C., Viallet, F., 2014. F0 characteristics in Parkinsonian speech: contrast between the effect of hypodopaminergy due to Parkinson's disease and that of the therapeutic delivery of L-Dopa. Revue de Laryngologie-otologie-rhinologie 135 (2), 63–70.

Gu, W., Fan, P., Liu, W., 2017, October. Acoustic analysis of mandarin speech in parkinson's disease with the effects of levodopa. In: International Seminar on Speech Production. Springer, Cham, pp. 211–224.

Guo, D., Yu, H., Hu, A., Ding, Y., 2016. Statistical analysis of acoustic characteristics of Tibetan Lhasa dialect speech emotion. In: SHS Web of Conferences, 25, p. 01017.

Hammen, V.L., Yorkston, K.M., 1996. Speech and pause characteristics following speech rate reduction in hypokinetic dysarthria. J. Commun. Disord. 29 (6), 429–445.

Hawley, M.S., Cunningham, S.P., Green, P.D., Enderby, P., Palmer, R., Sehgal, S., O'Neill, P., 2012. A voice-input voice-output communication aid for people with severe speech impairment. IEEE Trans. Neural Syst. Rehabil. Eng. 21 (1), 23–31.

Holmers, R., Oates, J., Phyland, D., Hughes, A., 2000. Voice characteristics in the progression of Parkinson's disease. Int. J. Lang. Commun. Disord. 35 (3), 407–418.

Hossain, N., Naznin, M., 2018. Sensing emotion from voice jitter. In: Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems, pp. 359–360.

Izard, C.E., 1971. The Face of Emotion. Appleton-Century-Crofts, East Norwalk, CT, US.

Illes, J., Metter, E.J., Hanson, W.R., Iritani, S., 1988. Language production in Parkinson's disease: acoustic and linguistic considerations. Brain Lang. 33 (1), 146–160.

Izard, C.E., 1994. Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psychol. Bull. 115, 288–299.

Jackson, P., & Haq, S. (2011, Apr). Surrey audio-visual expressed emotion (savee) database. University of Surrey: Guildford, UK. Retrieved from www.ee.surrey.ac.uk/Personal/P.Jackson/SAVEE/.

Johnstone, T., Scherer, K.R., 2000. Vocal communication of emotion. Handbook of emotions 2, 220–235.

Juslin, P., Scherer, K., Harrigan, J., Rosenthal, R., 2005. The New Handbook of Methods in Nonverbal Behavior Research. Oxford University Press, Oxford.

Kempler, D., Van Lancker, D., 2002. Effect of speech task on intelligibility in dysarthria: a case study of parkinson's disease. Brain Lang. 80 (3), 449–464.

Kim, Y., Lee, H., Provost, E.M., 2013. Deep learning for robust feature generation in audiovisual emotion recognition. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3687–3691.

Krom, G.D., 1995. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. J. Speech Lang. Hear. Res. 38 (4), 794–811.
Kumbhakarn, M., Sathe-Pathak, B., 2015. Analysis of emotional state of a person and its effect on speech features using praat software. In: 2015 International Conference on Computing Communication Control and Automation, pp. 763–767.
Laukka, P., Juslin, P., Bresin, R., 2005. A dimensional approach to vocal expression of emotion. Cogn. Emot. 19 (5), 633–653.
Le Dorze, G., Ouellet, L., Ryalls, J., 1994. Intonation and speech rate in dysarthric speech. J. Commun. Disord. 27 (1), 1–18.
Levenson, R.W., Ekman, P., Heider, K., Friesen, W.V., 1992. Emotion and autonomic nervous system activity in the minangkabau of west sumatra. J. Pers. Soc. Psychol. 62 (6), 972.
Li, X., Tao, J., Johnson, M.T., Soltis, J., Savage, A., Leong, K.M., Newman, J.D., 2007. Stress and emotion classification using jitter and shimmer features. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-icassp'07, 4 pp. IV–1081.
Liss, J.M., Weismer, G., 1994. Selected acoustic characteristics of contrastive stress production in control geriatric, apraxic, and ataxic dysarthric speakers. Clin. Linguist. Phon. 8 (1), 45–66.
Liu, L., Jian, M., Gu, W., 2019. Prosodic Characteristics of Mandarin Declarative and Interrogative Utterances in Parkinson's Disease. Age (year) 66 (6.00), 63–64.
Livingstone, S.R., Russo, F.A., 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE 13 (5), e0196391.
Ma, J.K., Hoffmann, R., 2010. Acoustic Analysis of Intonation in Parkinson's Disease. In: Eleventh Annual Conference of the International Speech Communication Association.
Ma, J.K.Y., Whitehill, T.L., So, S.Y.S., 2010. Intonation contrast in Cantonese speakers with hypokinetic dysarthria associated with Parkinson's disease. J. Speech Lang. Hear. Res.
McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., Nieto, O., 2015. librosa: audio and music signal analysis in python. In: Proceedings of the 14th Python in Science Conference, pp. 18–25.
Murray, I.R., Arnott, J.L., 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. J. Acoust. Soc. Am. 93 (2), 1097–1108.
Patel, R., 2002a. Phonatory control in adults with cerebral palsy and severe dysarthria. Augment. Alternative Commun. 18 (1), 2–10.
Patel, R., 2002b. Prosodic control in severe dysarthria. J. Speech Lang. Hear. Res.
Patel, R., 2003. Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy. J. Speech Lang. Hear. Res.
Patel, R., 2004. The acoustics of contrastive prosody in adults with cerebral palsy. J. Med. Speech Lang. Pathol. 12 (4), 189–193.
Plack, C.J., Barker, D., Hall, D.A., 2014. Pitch coding and pitch processing in the human brain. Hear. Res. 307, 53–64.
Platt, L.J., Andrews, G., Young, M., Quinn, P.T., 1980. Dysarthria of adult cerebral palsy: I. intelligibility and articulatory impairment. J. Speech Lang. Hear. Res. 23 (1), 28–40.
Rosen, K.M., Kent, R.D., Delaney, A.L., Duffy, J.R., 2006. Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. J. Speech Lang. Hear. Res.
Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E., 2011. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J. Acoust. Soc. Am. 129 (1), 350–367.
Scherer, K.R., 2003. Vocal communication of emotion: a review of research paradigms. Speech Commun. 40 (1–2), 227–256.
Scherer, U., Helfrich, H., Scherer, K.R., 1980. Paralinguistic behaviour: internal push or external pull? Language. Elsevier, pp. 279–282.
Schuller, B., Müller, R., Lang, M., Rigoll, G., 2005. Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles. In: Ninth European Conference on Speech Communication and Technology.
Schuller, B., Steidl, S., Batliner, A., 2009. The interspeech 2009 emotion challenge. In: Tenth Annual Conference of the International Speech Communication Association.
Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., 2013. The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: Proceedings interspeech 2013, 14th Annual Conference of the International Speech Communication Association. lyon, france.
Stanislavsky, K., Hapgood, E., Gielgud, J., 1936. An Actor Prepares. Theatre Arts. Inc., New York, ny.
Steinberg, S., 1995. Introduction to Communication Course Book 1: The basics (Vol. 1). Juta and Company Ltd.
Toivanen, J., Waaramaa, T., Alku, P., Laukkanen, A.-M., Seppänen, T., Väyrynen, E., Airas, M., 2006. Emotions in [a]: a perceptual and acoustic study. Logopedics Phoniatrics Vocol. 31 (1), 43–48.
Tomkins, S., 1962. Affect Imagery consciousness: Volume i: The positive Affects. Springer Publishing Company.
Tomkins, S., 1963. Affect Imagery consciousness: Volume ii: The negative Affects. Springer Publishing Company.
Van Bezooijen, R., Otto, S.A., Heenan, T.A., 1983. Recognition of vocal expressions of emotion: a three-nation study to identify universal characteristics. J. Cross Cult. Psychol. 14 (4), 387–406.
Walshe, M., Miller, N., 2011. Living with acquired dysarthria: the speaker's perspective. Disabil. Rehabil. 33 (3), 195–203.
Whiteside, S.P., 1998. Simulated emotions: an acoustic study of voice and perturbation measures. In: Fifth International Conference on Spoken Language Processing.
Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Deng, Z., Lee, S., Busso, C., 2004. An acoustic study of emotions expressed in speech. In: Eighth International Conference on Spoken Language Processing.