Speech Signal Analysis and Pattern Recognition in Diagnosis of Dysarthria

Minu George Thoppil1, C. Santhosh Kumar2, Anand Kumar1,2, John Amose3

1Department of Neurology, AIMS, Kochi, Kerala, 2Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham University, 3Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India

Abstract

Background: Dysarthria refers to a group of disorders resulting from disturbances in muscular control over the speech mechanism due to damage of central or peripheral nervous system. There is wide subjective variability in assessment of dysarthria between different clinicians. In our study, we tried to identify a pattern among types of dysarthria by acoustic analysis and to prevent intersubject variability. Objectives: (1) Pattern recognition among types of dysarthria with software tool and to compare with normal subjects. (2) To assess the severity of dysarthria with software tool. Materials and Methods: Speech of seventy subjects were recorded, both normal subjects and the dysarthric patients who attended the outpatient department/admitted in AIMS. Speech waveforms were analyzed using Praat and MATHLAB toolkit. The pitch contour, formant variation, and speech duration of the extracted graphs were analyzed. Results: Study population included 25 normal subjects and 45 dysarthric patients. Dysarthric subjects included 24 patients with extrapyramidal dysarthria, 14 cases of spastic dysarthria, and 7 cases of ataxic dysarthria. Analysis of pitch of the study population showed a specific pattern in each type. F0 jitter was found in spastic dysarthria, pitch break with ataxic dysarthria, and pitch monotonicity with extrapyramidal dysarthria. By pattern recognition, we identified 19 cases in which one or more recognized patterns coexisted. There was a significant correlation between the severity of dysarthria and formant range. Conclusions: Specific patterns were identified for types of dysarthria so that this software tool will help clinicians to identify the types of dysarthria in a better way and could prevent intersubject variability. We also assessed the severity of dysarthria by formant range. Mixed dysarthria can be more common than clinically expected.

Keywords: Ataxic, dysarthria, extrapyramidal, F0 jitter, formant range, pitch break, spastic

INTRODUCTION

Dysarthria refers to a group of speech disorders resulting from disturbances in muscular control over the speech mechanism due to damage of the central or peripheral nervous system.[1] Although there have been several attempts to improve speech recognition for dysarthric speakers, and other attempts to integrate articulatory knowledge into speech recognition, these efforts have not until recently converged. There is wide subjective variability in assessment of dysarthria between different clinicians. In our study, we tried to identify a pattern among types of dysarthria by pattern recognition and to see whether any acoustic parameter correlated with the clinical severity.

The Mayo Clinic classification of dysarthria includes six categories: (1) FLACCID, (2) spastic and “unilateral upper motor neuron (UMN),” (3) ataxic, (4) hypokinetic, (5) hyperkinetic, and (6) mixed dysarthria.[2] Speech is produced when air from the lungs is modulated by the vocal cord and vocal tract.[3] Dysarthric speech characteristics[4]

Darley et al. in 1075 described the acoustic quality of different types of dysarthria

1. Ataxic dysarthria which affects respiration, phonation, resonance, and articulation tend to place the same excessive stress on all syllables
2. Spastic dysarthria is characterized by the harshness of the vocal quality and long duration in phoneme to phoneme transitions and syllables. Pitch break can be seen

Address for correspondence: Dr. Minu George Thoppil,
Department of Neurology, Renal Medcity, Kochi, Kerala, India.
E-mail: minu.thomas82@gmail.com

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Thoppil MG, Kumar CS, Kumar A, Amose J. Speech signal analysis and pattern recognition in diagnosis of dysarthria. Ann Indian Acad Neurol 2017;20:352-7.
3. Hypokinetic dysarthria seen in Parkinson’s disease is characterized by hoarse speech with low volume and compulsive repetition of syllables with on monopitch and monoloudness.

4. Hyperkinetic dysarthria seen in Huntingtons disease is associated with harsh sounding, hypernasality, and frequent pauses. There is associated dystonia with lack of intelligibility.

5. Flaccid dysarthria due to lower motor neuron (LMN) paralysis of vocal cord shows harsh voice, low volume with inspiritional stridency.

6. Mixed dysarthria is characterized by harshness of voice in case of UMN involvement and breathy voice in case of LMN involvement.

Acoustic analysis of the speech can be done by fast Fourier transformation.[5]

**Pitch and formant frequency**

Air flowing through the glottis when measured as waveform, consist of three phases: closed phase, glottal open phase, and return phase. The time duration of one glottal cycle is referred to as the pitch period and the reciprocal of the pitch period is the corresponding pitch, which is also called as the fundamental frequency. Normal pitch range is about 60–400 Hz. Males have lower pitch than females because their vocal folds are longer and more massive. F0 jitter is a phenomenon by which pitch period vary over periods and is characteristic of harsh voice.[6] Formant frequency is first defined by Gunnar Fant in 1960,[7] as concentration of acoustic energy around a particular frequency in a speech wave. It is the spectral peaks of a sound spectrum.

The pitch or fundamental frequency is influenced by:

1. Vocal fold muscle tension - as the tension increases, the pitch increases
2. Vocal fold mass - as the mass increases, the pitch decreases because the folds are more sluggish
3. Air pressure behind the glottis in the lungs and trachea, which increase in a stressed sound or in a more excited state of speaking - as the pressure below the glottis increases, the pitch increases.

Normal pitch range is about 60–400 Hz. Males have lower pitch than females because their vocal folds are longer and more massive.

**Fourier transformation**

Fourier transformation is an operation that maps a function to its corresponding Fourier series or to an analogous continuous frequency distribution.[8] The Fourier transform decomposes any function into a sum of sinusoidal basis functions.[9]

**Acoustic characteristics of different types of dysarthria**

Pathophysiological changes early in the course of the parkinsonism can lead to changes in the ability of the central nervous system to control the musculature of the speech apparatus. This finding was most consistent in the reduced intonation in the early phases. F0 variability seen in parkinsons disease is seen during prodromal phase of illness can be used as a useful biomarker to evaluate the efficacy of pharmacological interventions in early disease process.[10] Formant analysis which is considered as a function of vocal tract can be affected by deficits in articulatory control and mobility of the same.[11] Zwirner and Barnes[12] reported increased variability of first formant (F1) values during vowel prolongations. Speakers with Parkinson’s disease (PD) were found to have reduced F1–F2 vowel space, compared to control speakers.[13] Connor et al.[14] reported that F1 and F2 transition rates were flatter in extrapyramidal dysarthria compared to control subjects. Flint et al.[15] examined F2 characteristics for PD and normal subjects and found flatter F2 transition rates in the PD patients. Le Dorze et al.[16] proposed smaller F0 difference in Parkinson patients compared to normal subjects. Canter[17] reported a higher F0 level and reduced F0 range in speech of patients with PD. Turner et al.[18] showed smaller vowel space areas in speech of amyotrophic lateral sclerosis patients compared with neurologically normal subjects. Ackermann suggested that increased pitch levels observed in dysarthric subjects may be related not to altered vocal tension but to altered sensory feedback from the laryngeal structures such that increased vocal effort is used by the ataxic speaker to overcome the sensory disturbance. They also noticed pronounced pitch fluctuations in the pitch contour among patients with ataxic dysarthria.[19]

In pseudobulbar palsy, there is apparent bimodal distribution in F0 range which is explained by different types of vocal manifestation in progressive bulbar palsy.[20] Canter[21,22] noted decreased F0 range during syllable production and during paragraph reading in parkinsons disease. Metter and Hanson[23] showed that there is decreased F0 variability in Parkinson’s disease compared to normal subjects.

Our aims of the study were:

1. Pattern recognition among types of dysarthria with software tool and to compare it with normal subjects
2. To compare the severity using clinical diagnosis and software tool.

**Materials and Methods**

**Institutional permission and study settings**

This study was approved by the Institutional Thesis Review Committee of Amrita Institute of Medical Sciences and Research Center. Consent for the participation was obtained from all the participants before the study. This study was performed in the Department of Neurology AIMS, Kochi, from January 2013 to December 2014.

**Study participant**

Primary literature search was performed using most frequently encountered keywords related with dysarthria and acoustic analysis. Given the absence of robust method of statistical sampling from literature, the sample size was arbitrarily estimated as “50.” Considering additional backup, seventy patients from hospital and neurology outpatient department were selected after consent for their speech analysis.
However, we excluded nonneurogenic causes of dysarthria and bed-ridden patients with dysarthria.

**Study design**

This was a noninterventional, cross-sectional comparative, observational study. The primary objective was to compare proportion of patients after both clinical diagnosis and acoustic analysis technique. The secondary objective was to confirm the difference in speech between normal subjects and dysarthric patients. The primary endpoint of this study was number of subjects diagnosed in all four types of dysarthria in both clinical diagnosis and machine learning group.

**Speech recording and speech-based dysarthria categorization**

Patients were asked to read one Malayalam (local language) paragraph and their speech was recorded using a Sony voice recorder (IC recorder-2GB intelligent noise cut recordable FM radio) under ideal conditions in a soundproof room in AIMS speech laboratory to avoid external sounds. The phrases “annan, ana, ela” and first sentence of the reading paragraph were extracted using audacity version 2.0.5, speech waveforms were exported to Praat vocal toolkit version (5.3.53) and pitch F0, formants F1 and F2 and pitch break were considered as deterministic parameters and were extracted for categorizing patients according to different types of dysarthria. Detailed study of extracted features to identify underlying characteristics within types of dysarthria was done using MATLAB toolkit version R2011b (7.13.0.564). Signal characteristics and nature of F0, F1, and F2 for each of four different types of dysarthria known from previously published literature were referred to categorize patients on basis of their disorder type. All patients were initially diagnosed by neurologist based on their phonations and later were subjected to acoustic analysis using software. Another neurologist blinded to previously made diagnosis performed recategorization of the patients based on findings from acoustic analysis.

**Statistical analysis**

To test the statistical significance of the association of clinical diagnosis with different categorical variables, Chi-square test was done. To compare the results of pattern recognition by software tool with the clinical diagnosis, McNemar Chi-square test was done. The validity parameters such as sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were computed for comparing assessment of pattern recognition by software tool with the clinical diagnosis. Severity-based classification was also done using formant range and by calculating F2 range/F1 range.

**Results**

Our study group included seventy persons. Of which 25 were normal subjects and 45 were those with dysarthria. The mean age of normal population was 53 and that of dysarthric population was 58 which was comparable. In our study population, there were 29 females and 41 males [Figure 5]. There were 25 normal subjects and 45 with dysarthria. Among the dysarthric group, there were 7 cases of ataxic dysarthria, 14 cases of spastic dysarthria, and 24 cases of extrapyramidal dysarthria. As per clinical severity, patients were divided into those with mild dysarthria and those with severe dysarthria. There were 24 case of mild dysarthria and 21 cases of severe dysarthria. We tried to identify specific patterns among types of dysarthria. Pitch was analyzed, it was found that F0 jitter is found to be associated with spastic dysarthria in 64.3% of cases and 25% of cases of extrapyramidal dysarthria. F0 jitter was found in 33% of cases of dysarthric subjects but was not found in any of the normal population. In ataxic dysarthria, pitch break was found in 6 out of 7 subjects. It was also found that the same phenomenon is present in only 28% of normal subjects but found in 56% of dysarthric population. When the extrapyramidal dysarthria was analyzed, it was found that F0 flat or monotonicity was found in 62.5% of extrapyramidal dysarthria, but only in 4% of spastic dysarthria and 57% of ataxic dysarthria. F0 flatness is found to be significantly associated with dysarthric patients but present only in 46.7% of normal population. The agreement of diagnosis by pattern recognition was compared with that of clinical diagnosis, it was found out that there is an accuracy of 62.7%. When the normal population and dysarthric population was compared on the basis of pattern recognition and clinical diagnosis, it was found that there is an accuracy of 85.7%, sensitivity of 93%, and specificity of 72%. It was also found that duration of speech in seconds increases as clinical severity increases. Formant range and F2/F1 range decrease as clinical severity increases.

**Discussion**

Acoustic analysis of normal and dysarthric population was done. The pitch and formant frequency of both were analyzed. Patterns recognized in each type of dysarthria are as follows:

- Spastic dysarthria - F0 jitter [Figure 1]
- Ataxic dysarthria - F0 break [Figure 2]
- Extrapyramidal dysarthria - F0 monotonicity [Figure 3].

F0 jitter or shimmer is a character described in pitch in which the pitch randomly varies over consecutive periods. The increased association of F0 jitter in dysarthric population may be explained by the harshness of the voice in this population which is due to time-varying characteristics of the vocal tract and vocal folds. Teager et al. reported that the character of F0 jitter is more associated with harsh speech. In a study conducted by Mori and Yasunori, it is described that F0 range of dysarthric speech is generally lower than that in normal population, and among the dysarthric group, this is more apparent in those with Parkinsonism. In a study conducted by Ackermann and Zeigler, they noticed F0 jitter above the normal range in 4 out of 11 subjects. Mavlov and Kehairov reported rapid modulations and oscillations of vocal amplitude as compared to normal subjects. Ackermann suggested that increased pitch levels observed in dysarthric subjects may be related not to altered vocal tension but to altered sensory feedback from the
laryngeal structures such that increased vocal effort is used by the ataxic speaker to overcome the sensory disturbance. They also noticed pronounced pitch fluctuations in the pitch contour among patients with ataxic dysarthria. Mori and Yasunori[24] reported that F0 range will be less for extrapyramidal dysarthria compared to normal population. This finding correlated with the study conducted by us.

However, more than one pattern was identified in 19 patients. It could be possible that these patients had mixed dysarthria by pattern recognition although clinically there appeared to be pure spastic or extrapyramidal dysarthria.

In our study, it was found that formant range and F2/F1 range decrease when severity increases [Figure 4]. Connor et al.[18] found that F1 and F2 transition rates were flatter in extrapyramidal dysarthria compared to control subjects. Flint et al.[15] examined F2 characteristics for PD and normal subjects and found flatter F2 transition rates in the PD patients during sentence reading. We calculated the duration of speech and found that as duration of speech increases, clinical severity also increases.

**Conclusions**

1. Different types of dysarthria when analyzed with software tool extracting pitch and formants showed specific patterns which correlated with the clinical diagnosis and could help prevent intersubject variability
2. Software tool can be used to assess the severity of dysarthria and hence could also provide ground for developing home-based biofeedback program
3. Mixed dysarthrias can be more common than clinically suspected.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

References
1. Gonzalez-Moreira E, Torres D, Ferrer CA, Ruiz Y. Improving dysarthria classification by pattern recognition techniques based on a bionic model. In: Ruiz-Shulcloper J, di Baja GS, editors. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Internet: Springer Berlin Heidelberg; 2013. p. 246-53.
2. Bradley WG, Daroff RB, Fenichel GM, Jankovic J. Neurology in Clinical Practice. 6th ed. Elsevier Saunders; 2012. p. 161-2.
3. Barney A, Shadle CH, Davies PO. Fluid flow in a dynamical mechanical model of the vocal folds and tract. 1: Measurements and theory. J Acoust Soc Am 1999;105:444-55.
4. Selouani SA, Dahmani H, Amami R, Hamam H. Using speech rhythm knowledge to improve dysarthric speech recognition. Int J Speech Technol 2012;15:57-64.
5. Kent RD, Weismer G, Kent JF, Vorperian HK, Duffy JR. Acoustic studies of dysarthric speech: Methods, progress, and potential. J Commun Disord. 1999;32:141-80, 183.
6. Teager HM, Teager SM. Evidence for nonlinear sound production mechanisms in the vocal tract. In: Hardcastle WI, Marchal A, editors. Speech Production and Speech Modeling. NATO Advanced Study Institute Series D, Bonas, France. Vol. 55. Boston, MA: Kluwer Academic Publishers; 1990. p. 241-61.
7. Atal BS, Hanauer SL. Speech analysis and synthesis by linear prediction of the speech wave. J Acoust Soc Am 1971;50:637-55.
8. Fourier Transform-definition of Fourier Transform by The Free Dictionary.
9. Quatieri TF. Discrete time speech signal processing. Principles and Practice.Chapter1.1. Discrete time speech signal processing.Library of congress Publication.2002. p. 10-11.
10. Harel BT. Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment. J Neurolinguistics 2004;6:439-53.
11. Goberman AM, Coelho C. Acoustic analysis of parkinsonian speech I: Speech characteristics and L-dopa therapy. NeuroRehabilitation 2002;17:237-46.
12. Zwirner P, Barnes GJ. Vocal tract steadiness: A measure of phonatory and upper airway motor control during phonation in dysarthria. J Speech Hear Res 1992;35:761-8.
13. Weismer G, Jeng JY, Laures JS, Kent RD, Kent JF. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. Folia Phoniatr Logop 2001;53:1-18.
14. Connor NP, Ludlow CL, Schulz GM. Stop consonant production in isolated and repeated syllables in Parkinson’s disease. Neuropsychologia 1989;27:829-38.
15. Flint AJ, Black SE, Campbell-Taylor I, Gailey GF, Levinton C. Acoustic analysis in the differentiation of Parkinson’s disease and major depression. J Psycholinguist Res 1992;21:383-9.
16. Le Dorze G, Ryalls J, Brassard C, Boulanger N, Ratté D. A comparison of the prosodic characteristics of the speech of people with Parkinson’s disease and Friedreich’s ataxia with neurologically normal speakers. Folia Phoniatr Logop 1998;50:1-9.
17. Canter GJ. Speech characteristics of patients with Parkinson’s disease: 1.
Intensity, pitch, and duration. J Speech Hear Disord 1963;28:221-9.
18. Turner GS, Tjaden K, Weismer G. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. J Speech Hear Res 1995;38:1001-13.
19. Murdoch BE. Ataxic dysarthria clinical features. Dysarthria: A Physiological Approach to Assessment and Treatment. Ch. 8. University of Wisconsin-Madison, Madison, WI, USA: Stanley Thornes Publishers; 1998. p. 250.
20. Hirose H, Imaizumi S, Yamori M. Voice quality in patients with neurological disorders. In: Fujimura O, Hirano M, editors. Voice Quality Control. San Diego: Singular; 1995. p. 235-48.
21. Canter G. Speech characteristics of patients with Parkinson’s disease: I. Intensity, pitch, and duration. J Speech Hear Disord 1963;28:221-9.
22. Canter G. Speech characteristics of patients with Parkinson’s disease: II. Physiological support for speech. J Speech Hear Dis 1965;30:44-9.
23. Metter EJ, Hanson WR. Clinical and acoustical variability in hypokinetic dysarthria. J Commun Disord 1986;19:347-66.
24. Mori H, Kobayashi Y. F0 and formant frequency distribution of dysarthric speech: A comparative study. Published by Acoustical society of America; 2004.