Acoustic Features of Dysphonic Speech vs Normal Speech in New Zealand English Speakers

Maryam Erfanian Sabaee
School of Computing
Unitec Institute of Technology
Auckland, New Zealand
merfanian@unitec.ac.nz

Hamid Sharifzadeh
School of Computing
Unitec Institute of Technology
Auckland, New Zealand
hsharifzadeh@unitec.ac.nz

ABSTRACT

This poster presents the acoustic features of distorted speech in three dysphonic New Zealanders compared to three healthy individuals as the control group. There are a total of six subjects in this study. Voice onset time and measurements of the two formant frequencies of vowel articulation in /hVd/ words are calculated for both groups, and the results are provided for comparison purposes. Furthermore, independent sample t-tests are conducted between dysphonic and normal speech results to evaluate the significancy of the differences in acoustic measurements.

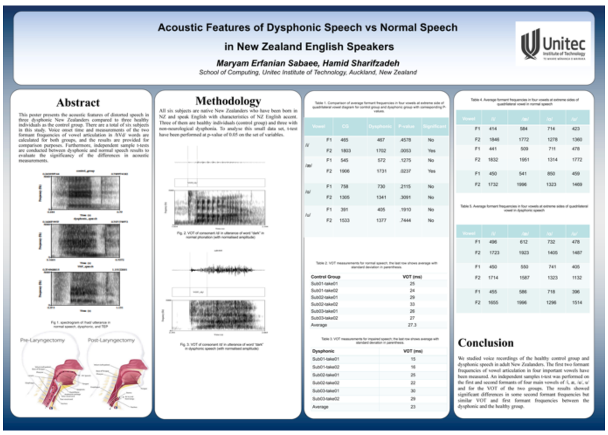
Keywords: Acoustic features, formants, distorted speech, dysphonia.

1. INTRODUCTION

In normal speech, voiced phonemes are generated through periodic vibration of vocal folds, which produces glottal harmonic airflow into the different chambers of the upper vocal tract. The utterance is initiated by exhalation, which induces a stream of air through the trachea and larynx, exiting via the oral and nasal chambers [1].

Some neurological and non-neurological diseases may cause speech impairments. Motor speech disorders (MSDs) are caused by nervous system diseases (neurological in origin). Dysarthria and apraxia are both considered to be kinds of MSDs [2]. Neurological speech impairment is not the focus of this paper.

The non-neurological collapse of speech can be classified into dysphonia and musculoskeletal disorders. Dysphonia can be caused due to various reasons such as inflammation, vocal cord paralysis, vocal nodules and polyps, scarring in vocal cords due to trauma or radiation, and even psychogenic cause such as depression or schizophrenia. Musculoskeletal disorders refer to abnormalities in bone and cartilage in the cleft and palate or any injuries in the musculoskeletal system [3].



Although voice disorders are not usually life-threatening, they have a profound effect on people's day-to-day lives. Therefore, significant research efforts have been made in recent years to understand distorted speech characteristics [4]. Acoustic features of distorted speech provide fundamental data required to develop speech reconstruction, speech recognition, and speech enhancement systems, which ultimately aim to improve the quality of living for these individuals [5, 6]. By analysing formant frequencies and voice onset time (VOT) of distorted speech (dysphonia) and comparing them to normal speech in New Zealand English, we aim to add to existing knowledge. Formants are very distinguishable frequency components of the acoustic signals produced by speech and are defined as resonant frequencies of the vocal tract. Temporal measures can also provide valuable information, particularly on pathologic voices. VOT is one of the temporal measurements which defines the length of time between the onset of the articulatory stop release burst (release of stop consonant) and the first glottal pulse of the following vowel [7, 8].

2. MATERIAL AND METHODS

All six subjects are native New Zealanders who have been born in NZ and speak English with characteristics of NZ English accent [9]. Three of them are healthy individuals (control group) and three with non-neurological dysphonia. In the dysphonic group, the clinical details are as follows: In Sub01, the source of the voice disorder was paralysed left larynx, Sub02 suffered from spasmodic dysphonia (the muscles in the larynx go into periods of contraction), and Sub03 underwent an organ-preserving operation to remove nodules on the vocal cords followed by radiation therapy which caused the formation of webs on his vocal cords. Therefore, apart from having some complications, the overall upper vocal tract in all three dysphonic participants has the same anatomy as a normal larynx. Audio recordings were made of subjects articulating /hVd/ words containing four vowels at extreme sides of quadrilateral vowel diagram (/i, æ, ɑ, u/). Praat was used to observe the formant frequencies from voice signal spectrograms and VOT calculation. The results were manually verified with MATLAB's power spectral density graph as described in [10]. The stop consonant (letter /d/ in "dark") to vowel transitional change (d- vowel) was chosen for the purpose of this study. T-test has been performed at the p-value of 0.05 on the set of variables.

3. RESULTS AND DISCUSSION

Table 1 presents the average of formant frequencies from two recordings in four vowels (/i, æ, ɑ, u/) at extreme sides of the quadrilateral in the control group. The result of these measurements is consistent with the literature.

Notably, the formant frequencies in open-front vowels /æ/ in NZ English are shifted compared to standard Received Pronunciation (RP). Table 2 outlines the average of formant frequencies from two recordings of vowels (/i, æ, ɑ, u/) in dysphonic speech. Also, the formant frequencies are mainly in a close range for all three subjects.

Table 1: Average formant frequencies (from two recordings) in four vowels at extreme sides of quadrilateral vowel diagram in normal speech.

Sub01	/i/	/æ/	/ɑ/	/u/
F1	414	584	714	423
F2	1846	1772	1278	1360
Sub02				
F1	441	509	711	478
F2	1832	1951	1314	1772
Sub03				
F1	450	541	850	459
F2	1732	1996	1323	1469

Table 2: Average formant frequencies (from two recordings) in four vowels at extreme sides of quadrilateral vowel diagram in dysphonic speech.

Sub01	/i/	/æ/	/ɑ/	/u/
F1	496	612	732	478
F2	1723	1923	1405	1487
Sub02				
F1	450	550	741	405
F2	1714	1587	1323	1132
Sub03				
F1	455	586	718	396
F2	1655	1684	1296	1514

Table 3: Comparison of average voice onset time for the control and dysphonic groups with corresponding p-values.

VOT	Control	Dysphonia	P-value	Significant
/d/ in dark	27.3	23	0.9	No

Table 3 shows the average VOT of two takes for three control group subjects and dysphonic speech, respectively.

In Table 4, the p-values confirm that the differences between the first formants in all vowels are not significant. However, the results for the second formant in vowel /i/ and /æ/ are significantly different. We can see that in most vowels, the results of dysphonic speech are generally consistent with the control group (i.e. no significant differences). However, the voice signal has a slight upward shift in the first formant, and the hoarseness of the voice reduces the periodic form of the signal in the spectrogram.

Table 4: Comparison of average formant frequencies in four vowels at extreme sides of the quadrilateral vowel diagram for the control and dysphonic groups with corresponding p-values.

Vowel		Control	Dysphonic	P-value	Significant
/i/	F1	465	467	.45788	No
	F2	1803	1702	.00539	Yes
/æ/	F1	545	572	.12753	No
	F2	1906	1731	.02378	Yes
/ɑ/	F1	758	730	0.2152	No
	F2	1305	1341	0.3091	No
/u/	F1	391	405	0.1910	No
	F2	1533	1377	0.7444	No

4. CONCLUSION

We studied voice recordings of the healthy control group and dysphonic speech in adult New Zealanders. The first two formant frequencies of vowel articulation in four important vowels have been measured. In addition, an independent samples t-test was performed on the first and second formants of four main vowels of /i, æ, /ɑ/, u/ and for the VOT of the two groups. The results showed significant differences in some second formant frequencies but similar VOT and first formant frequencies between the dysphonic and the healthy group.

5. REFERENCES

[1] G. Fant, "Acoustic theory of speech production," Mouton, The Hague,1960.

[2] J. R. Duffy, "Motor Speech Disorders Substrates, Differential Diagnosis, and Management," Elsevier - Health Sciences Division, United States, 2012.

[3] E. D. Ross and A. J. Rush, "Diagnosis and neuroanatomical correlates of depression in brain-damaged patients," Journal of Archives of General Psychiatry, vol. 38, pp. 1338–1344, 1981.

[4] I. V. McLoughlin, O. Perrotin, H. Sharifzadeh, J. Allen, and Y. Song, "Automated assessment of glottal dysfunction through unified acoustic voice analysis," Journal of Voice, In Press, no. https://doi.org/10.1016/j.jvoice.2020.08.032, 2020.

[5] H. Sharifzadeh, I. McLoughlin and F. Ahmadi, "Reconstruction of Normal Sounding Speech for Laryngectomy Patients Through a Modified CELP Codec," in IEEE Transactions on Biomedical Engineering, vol. 57, no. 10, pp. 2448-2458, 2010.

[6] I. V. Mcloughlin, H. R. Sharifzadeh, and S. Tan, "Reconstruction of phonated speech from whispers using formant-derived plausible pitch modulation," ACM Transactions on Accessible Computing, vol. 6, pp. 121, 2015.

[7] M. E. Sabaee, H. Sharifzadeh, I. Ardekani, and J. Allen, "A preliminary acoustic analysis of laryngectomised speech in adult New Zealanders," 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 271–275, 2018.

[8] H. Sharifzadeh, I. V. McLoughlin, and M. J. Russell, "A comprehensive vowel space for whispered speech," Journal of Voice, vol. 26, pp. 49–56, 2012.

[9] C. Watson, M. Maclagan, and J. Harrington, "Acoustic evidence for vowel change in New Zealand English," Language Variation and Change, vol. 12, no. 1, pp. 51–68, 2000.

[10] K. Mustafa and I. Bruce, "Robust formant tracking for continuous speech with speaker variability," IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, pp. 435 – 444, 2006.