Phonatory Analysis on Parkinson's Disease Patients Attending Singing and Discussion Therapy (Parkinsonics) using Signal Processing Techniques

C. Chen[1], L. Moro-Velazquez[1], A. S. Ozbolt[1], A. Butala[2,3], A. Pantelyat[2] and N. Dehak[1]

1. Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland, USA
2. Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
3. Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
{cchen247, laureano, aozbolt1, ndehak3}@jhu.edu, {Ankur.Butala, apantel1}@jhmi.edu

***Abstract*— Persons with Parkinson's Disease (PD) frequently have speech and voice disorders. Regular speech therapy with a speech-language pathologist is essential to mitigate progressive symptom deterioration. Speech-related therapies, such choral singing groups are alternative approaches designed to be more naturalistic and enhance participant enjoyment. It is important to measure and quantify the effects of these therapies on the vocal features of PD patients to determine efficacy. We performed a prospective crossover study of 25 PD patients attending discussion or choral-singing groups for 12 weeks each (Parkinsonics NCT02753621). Every six weeks, each participant produced several recordings of the sustained vowels /a:/ and /e:/ at 'normal' and 'maximum' loudness. The goal was to identify if there are signal-processing-based features that can help track changes in the voice of PD patients over time. Voice features were extracted from these recordings using the Automatic Voice Condition Analysis (AVCA) library and were compared using non-parametric statistical tests. Results suggest that neither therapy caused any significant improvements in the analyzed phonatory aspects of the patients' voices. Future work should require use of connected speech to analyze articulation and comparison with a control group of participants with PD not attending any therapy to evaluate if therapy can mitigate the progressive effects of PD on the voice of patients.**

Keywords*— *Parkinson's Disease, phonation, sustained vowels, statistical analysis

I. Introduction

Parkinson's Disease (PD) is a neurological condition that eventually leads to motor impairments, including visual and speech dysfunctions [1]. These neurodegenerative processes can cause both dysphonia and dysarthria [2]. Dysphonia is the inability to produce a normal phonation caused by the dysfunction of the phonatory system, whereas dysarthria is characterized by voice and speech impairments that impact a person's ability to communicate effectively [3]. Some of the symptoms resulting from dysarthria include monoloudness and pitch, reduced loudness, harshness, and breathiness [3], [4]. Furthermore, the laryngeal and respiratory dysfunction caused by PD can also lead to voice problems. PD patients often have softer voices because it takes greater respiratory effort to create a similar level of loudness to healthy people because they have increased laryngeal resistance to air flow and decreased expiratory power [5]. In order to relieve some of these symptoms, standardized speech therapies have been developed such as the Lee Silverman Vocal Training (LSVT) LOUD program. However, this requires close coordination with a certified speech and language pathologist, hence restricting accessibility. Therefore, other alternatives were developed such as choral singing therapy. Singing can be used as a form of therapy because singing uses the larynx as the primary source of sound and the respiratory system for initiation of the vocal folds [6]. Singing therapy can help improve intonation, timing, speech rates, and respiratory capabilities as singing involves range, pitch variability, and different tempos [7]. Another form of therapy used to help PD patients is speech therapy. This type of therapy has been found to improve speech intelligibility and different prosodic aspects of speech such as vocal loudness and pitch [8].

The goal of this paper was to identify if there are signal-processing-based features that can help track changes in the voice of PD patients along time and to analyze how these features change with therapy and the advance of the disease. This was accomplished through a prospective and crossover study of 25 PD patients who attended both singing and discussion therapy.

II. Related work

There have been several studies that suggest the influence of PD on the articulatory and phonatory systems. In one specific study [9], the authors recorded samples for 39 PD patients and 60 non-PD patients of sustained vowel phonations. A significant difference between classes was found regarding noise and tremor related measurements. Many other studies have focused only on the phonatory aspect and employed one or more sustained vowels during experimental trials in order to extract voice signals that were used to perform a voice quality analysis [4].

Features such as jitter, shimmer, and noise, or Harmonic to Noise Ratio (HNR) or Noise to Harmonic Ratio (NHR), are commonly used to quantify the effects of Parkinson's. Multiple studies [9]–[11] that have analyzed recordings of sustained vowel sounds for both patients and controls concluded that jitter and shimmer are higher

and HNR is lower in PD patients. However, other studies have observed conflicting results. In one study [3], it was found that jitter was only prominent in late-stage Parkinson's patients, while shimmer and NHR had no significant statistical difference when comparing patients and controls. Another study [12] found that shimmer and Parkinson's had a higher correlation than jitter.

There have also been studies that performed longitudinal analyses on voice quality of Parkinson's patients. One study [13] evaluated jitter, shimmer, and NHR in sustained vowel recordings from 80 PD patients and 60 controls taken over two sessions which were on average 33 months apart for the patients and 25 months apart for the controls. The authors found that shimmer and NHR in both males and females with PD had significantly higher values than the controls, while there was no statistical difference in jitter between the groups. Another study [14] measured 19 patients before and after they were treated with symptomatic medicine such as levodopa or dopamine agonist. They analyzed jitter, shimmer, HNR, Recurrence Period Density Entropy (RPDE), Maximum Phonation Time (MPT), and Pitch Period Entropy (PPE). There were significant differences between patients and controls in each measure except for RPDE and MPT. Both of these studies suggest that the treatment tends to improve voice quality in only certain aspects.

Fewer studies have compared the difference in the voices of PD patients over multiple sessions of therapy. In one [15], patients attended speech and choral singing therapy for a collective 20 hours of speech and 26 hours of singing therapy over a five month period. The authors found that there was a significant improvement in the maximum duration of sustained vowel phonation. Another study found that the maximal phonation times for sustained vowels increased following 13 group speech therapy sessions in a one month [16]. This indicates that over time, PD patients' phonatory capacity tends to improve with the aid of therapy.

III. Materials

The Parkinsonics data set contains audio recordings of 25 patients with PD who fully completed both the speech and singing therapy. The patients were recruited from several places such as private practices and University-based Movement Disorders clinics. Each participant signed an informed consent form that was approved by the Johns Hopkins Medical Institutions Institutional Review Board. The corpus was randomly divided into two subsets: group 1 attended weekly singing therapy and group 2 attended weekly discussion therapy. After 12 weeks, the groups swapped their therapeutic strategies. People in group 1 attended weekly discussion therapy and group 2 started weekly singing therapy. Recordings were taken before the participants started therapy and continued every six weeks for 30 weeks. The session timeline appears in Figure 1. Each recording session involved patients voicing the vowel sound /a:/ and /e:/ for as a long as possible in one breath. For each vowel, three trials were conducted at high intensity, which in this study are referred to as A loud and E loud, and three trials were conducted at a normal speaking level, which are referred to as A norm and E norm.

IV. Methods

Features were extracted from the Parkinsonics data set sustained vowels using the Automatic Voice Condition Analysis (AVCA) library [17]. The AVCA library produces 261 coefficients per recording that represent the average and standard deviation of four voice feature families: amplitude and frequency perturbation and fluctuation, spectral-cepstral, complexity, and modulation spectra. However, the spectral-cepstral features, which include mel-frequency cepstral coefficients (MFCC), perceptual linear predictive coefficients (PLP), and Modulation Spectra Centroids (MSCents) and dynamic range (MSDR), were not considered because they are difficult to physically interpret, which left each recording with 69 calculated features, including statistics such as mean and standard deviation for several of them. Only features that can be physically interpreted have been included to measure the patients' potential changes in their voice. An overview of these features can be found in [17], [18] and in Table 1. These were normalized by subtracting the mean and dividing by the standard deviation. The means and standard deviations were calculated per feature considering all six sessions together. Every participant had 12 feature vectors of dimension 69 that represented each one of the 3 trials of A norm, A loud, E norm, and E loud for each session.

Table 1. AVCA library features employed in this study

Feature family	**Coefficients**
Amplitude and frequency perturbation and fluctuation	Absolute and relative jitter and shimmer, RAP, PPQ5, APQ3, APQ5, FTRI, ATRI, and statistics about HNR, NHR, CHNR, NNE, and GNE
Spectral-Cepstral	LHr
Complexity	D2, LZC, and statistics about LLE, ApEn, SampEn, GSampEn, FuzzyEn, mSampEn, PE, RPDE, and DFA
Modulation Spectra	MSP, and statistics about MSH, MSW, CIL, RALA, and LMR

To measure the changes in each of the coefficients as the patients attended therapy, each feature vector of session 1 (before any therapy) was compared to the respective feature vector of session 6 (at the end of the therapy period) using the Kruskal-Wallis and the Wilcoxon rank-sum tests. These both measure whether the samples originate

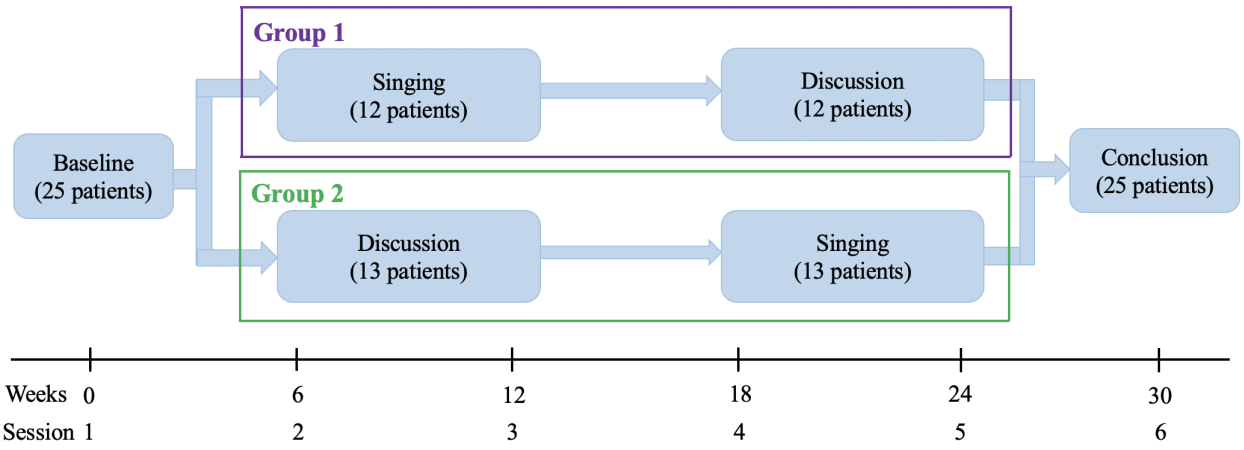

Figure 1. Timeline of Therapy Sessions for Groups 1 and 2

from the same distribution and by comparing the group medians. A significant Kruskal-Wallis or Wilcoxon rank-sum test yielding $p < 0.05$ respectively indicates that the group medians are not equal or the groups do not have the same distribution. We selected non-parametric tests because the distributions of the Parkinsonics coefficients are not normal. After running these experiments, we applied the false discovery rate (FDR) correction for multiple comparisons [19] to the p-values. For each of the significant features, the slopes of the features' progression along the six sessions were calculated using linear regression. A negative slope will indicate that the feature tends to decrease and vice-versa. We also studied whether the time since medication (TSM)[1] given to the patients is an underlying explanation for significant vocal feature changes. Therefore, we calculated the correlation between the changes in TSM between session 1 and 6 and the changes in the vocal feature values between the same sessions. If there is a high correlation between changes in TSM and vocal features, that might be motivated by the changes in TSM rather than the therapy or the advance of the disease.

Groups 1 and 2 were studied separately because they received the therapy differently (as explained in Figure 1). Additionally, we did not study the differences between the two groups because the age distribution of group 1 differs from the age distribution of group 2, as shown in Figure 2. The original groups of this corpus were larger and age-matched, but some of the participants dropped out before completing the study; resulting in the mismatched age distributions indicated in the mentioned figure.

[1]The time since medication in this study refers to the elapsed time between the PD-related medicine intake and time of recording

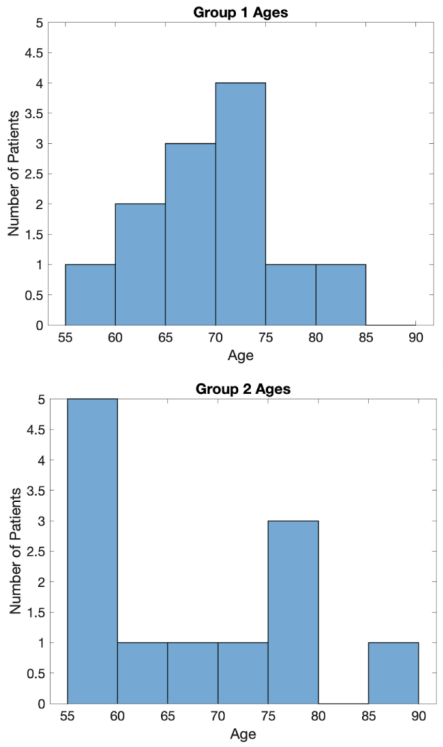

Figure 2. Age distributions of Groups 1 and 2

V. Results and Discussion

Table 2 includes the comparison of group 1's feature values from session 1 to 6, while Table 3 contains the comparison of group 2's coefficient values from session 1 to 6. The coefficient columns of these tables list the significant features according to both the Kruskal-Wallis and Wilcoxon rank-sum test following the FDR correction.

The average change column shows the average difference

of patients' vocal feature values between session 6 and session 1. A negative average change indicates that the coefficient value decreased in session 6 compared to session 1. The standard deviation column is the standard deviation of the average change. The slope column indicates the slope of each patients' linear regression line of their average coefficient values for the six sessions. This column has been included to confirm if the trend of each feature across the six sessions is consistent with the analysis before and after therapy. An example of how this slope was calculated can be seen in Figure 3.

Table 2. Group 1 Sessions 1 to 6

Coefficient	Average Change	Standard Deviation	Slope
A Norm			
-	-	-	-
A Loud			
CHNR_std	-1.08	0.95	-0.18
CHNR_mean	-1.03	0.92	-0.17
E Norm			
CHNR_std	-0.9	1.39	-0.19
CHNR_mean	-0.82	1.36	-0.17
E Loud			
CHNR_std	-1.13	1.12	-0.24
CHNR_mean	-1.09	1.03	-0.22
GNE_mean	-0.26	1.46	-0.067

All of these features are statistically significant.

Table 3. Group 2 Sessions 1 to 6

Coefficient	Average Change	Standard Deviation	Slope
A Norm			
LMR_std*	-0.82	1.15	-0.087
A Loud			
-	-	-	-
E Norm			
LMR_std	-1.23	0.99	-0.15
GNE_mean	-0.79	1.26	0.0027
CHNR_std	-0.74	0.92	-0.15
MSHmod_mean	-0.65	1.31	-0.094
E Loud			
CHNR_std	-1.03	1.20	-0.17
CHNR_mean	-0.94	1.07	-0.15
GNE_std	0.82	1.22	0.14
GNE_mean	-0.77	1.09	0.056
LMR_std	-0.81	0.90	-0.071

All of these features are statistically significant.
*Negative correlation of $\leq$ -0.5 between change in TSM and change in feature values

There are several significant features that can be seen across A norm, A loud, E norm, and E loud when comparing sessions 1 and 6 for groups 1 and 2. The cepstral-harmonics-to-noise-ratio (CHNR) mean and standard deviation decreased for A loud, E norm, and E loud in group 1. The CHNR mean only decreased for E loud, and the standard deviation decreased for E norm, and E loud in group 2. CHNR measures the difference in the levels of the speech harmonics and noise cepstrum. In a typical voice, the noise cepstrum magnitude should be

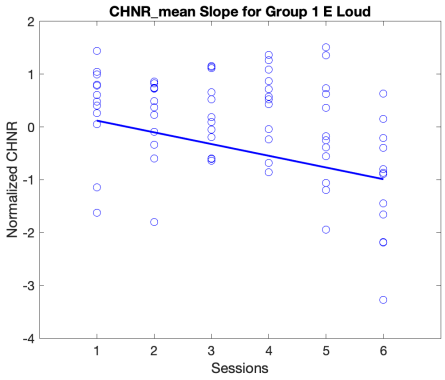

Figure 3. Slope of CHNR mean for E Loud Group 1. This figure depicts how the mean CHNR decreases with time for Group 1.

less than that of the harmonics cepstrum. The therapy did not help improve CHNR as the mean value tends to decrease, as indicated also in Figure 3; this indicates that on average, the patients' speech was more noisy in session 6 than in session 1.

The glottal-to-noise excitation ratio (GNE) mean decreased for E loud in group 1 and for E norm and E loud in group 2. The GNE describes the aspiration that is present in voice. Aspiration is the audible breath that accompanies voice and speech. Since the mean decreased, for both group 1 and 2, the patients experienced more audible breath and noise in their voices after session 6.

The low modulation ratio (LMR) standard deviation tends to be reduced for group 2, indicating a reduction of pitch tremor variability. In the same manner, Modulation Spectrum Homogeneity (MSHmod) also decreases with time, indicating that the modulation spectrum of the voice of the speakers is less homogeneous with time, and therefore, noisier.

For group 2, the LMR std in A norm had a negative correlation lower than -0.5 between the difference in TSM and the difference in feature values. The fact that the changes in this voice feature correlates with the changes in the time between medicine intake and recording might suggest that this voice change could be motivated by the differences in medicine intake instead of because of the advance of PD or the effects of the therapy.

In general, these results suggest that therapy did not have a significant impact in most of the vocal features included in Table I. Only features measuring vocal noise turned out to be significant when comparing pre- and post-therapy recordings, and these indicate that the voice of participants tends to be noisier with time. One reason could be that the advance of the disease affects phonation in this sense, and features such as CHNR, GNE and MSHmod would be sensitive to the changes in the noise

of the voice of patients. However, a time span of 30 weeks might be small to allow big changes in the voice of patients caused by disease progression. There can be multiple reasons why only the noise-related features change. The first one could be that these two therapies do not have a particular effect in the phonatory capabilities of participants. The second is that these features or the acoustic materials employed (sustained vowels) do not capture certain phonatory changes, or other aspects such as loudness, prosody, or articulation, that might be affected by therapy.

VI. Conclusions and future work

In this paper, we studied the effects of discussion and singing therapy on a set of signal processing features indicated for the analysis of voice conditions, employing the voices of 25 PD patients. This was done by extracting features from the recordings of the sustained vowels /a:/ and /e:/ at normal and maximum loudness using the AVCA feature library. The goal of the study was twofold: to identify if there are signal-processing-based features that can help track changes in the voice of PD patients along time and to analyze how these features change along time with the therapy and the advance of the disease. The features at the beginning and end of the study (30 weeks time-span) were compared using the Kruskal-Wallis test and the Wilcoxon rank-sum test, and all of the statistically significant features were reported. There were several significant features that were found across both phonations and loudness levels that indicate that there was an increase in the amount of noise and breathiness in the patients' speech. From this, we can conclude that discussion and singing therapy did not improve these quality of voice measurements. Another possible interpretation is that the features used in this study are not sensitive to other potential changes in the participant's phonatory capabilities. In any case, we cannot conclude that therapy is completely ineffective because there was no control group of either people without PD or people with PD who did not attend therapy that was also recorded and analyzed.

In future work, additional speech-related features should be employed to analyze if the therapy had an impact in the articulation of the participants. For those tasks we will use connected speech instead of sustained vowels.

References

[1] J. Kegl, H. Cohen, and H. Poizner, "Articulatory consequences of parkinson's disease: Perspectives from two modalities," *Brain and Cognition*, vol. 40, no. 2, pp. 355–386, 1999.

[2] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of Speech and Hearing Research*, vol. 12, no. 2, pp. 246–269, 1969.

[3] R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, "Voice characteristics in the progression of parkinson's disease," *International Journal of Language & Communication Disorders*, vol. 35, no. 3, pp. 407–418, 2000.

[4] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, "Advances in parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects," *Biomedical Signal Processing and Control*, vol. 66, p. 102418, 2021.

[5] B. J. De Swart, S. Willemse, B. Maassen, and M. Horstink, "Improvement of voicing in patients with parkinson's disease by speech therapy," *Neurology*, vol. 60, no. 3, pp. 498–500, 2003.

[6] E. L. Stegemöller, H. Radig, P. Hibbing, J. Wingate, and C. Sapienza, "Effects of singing on voice, respiratory control and quality of life in persons with parkinson's disease," *Disability and Rehabilitation*, vol. 39, no. 6, pp. 594–600, 2017.

[7] L. C. Shih, J. Piel, A. Warren, L. Kraics, A. Silver, V. Vanderhorst, D. K. Simon, and D. Tarsy, "Singing in groups for parkinson's disease (sing-pd): A pilot study of group singing therapy for pd-related voice/speech disorders," *Parkinsonism & Related Disorders*, vol. 18, no. 5, pp. 548–552, 2012.

[8] G. M. Schulz and M. K. Grant, "Effects of speech therapy and pharmacologic and surgical treatments on voice and speech in parkinson's disease: A review of the literature," *Journal of Communication Disorders*, vol. 33, no. 1, pp. 59–88, 2000.

[9] Y. Tanaka, M. Nishio, and S. Niimi, "Vocal acoustic characteristics of patients with parkinson's disease," *Folia Phoniatrica et Logopaedica*, vol. 63, no. 5, pp. 223–230, 2011.

[10] F. J. Jiménez-Jiménez, J. Gamboa, A. Nieto, J. Guerrero, M. Orti-Pareja, J. A. Molina, E. García-Albea, and I. Cobeta, "Acoustic voice analysis in untreated patients with parkinson's disease," *Parkinsonism & Related Disorders*, vol. 3, no. 2, pp. 111–116, 1997.

[11] Y.-I. Bang, K. Min, Y. H. Sohn, and S.-R. Cho, "Acoustic characteristics of vowel sounds in patients with parkinson disease," *NeuroRehabilitation*, vol. 32, no. 3, pp. 649–654, 2013.

[12] M. Novotnỳ, P. Dušek, I. Daly, E. Ržička, and J. Rusz, "Glottal source analysis of voice deficits in newly diagnosed drug-naïve patients with parkinson's disease: Correlation between acoustic speech characteristics and non-speech motor performance," *Biomedical Signal Processing and Control*, vol. 57, p. 101818, 2020.

[13] S. Skodda, W. Grönheit, N. Mancinelli, and U. Schlegel, "Progression of voice and speech impairment in the course of parkinson's disease: A longitudinal study," *Parkinson's Disease*, vol. 2013, 2013.

[14] J. Rusz, R. Čmejla, H. Ržičková, J. Klempíř, V. Majerová, J. Picmausová, J. Roth, and E. Ržička, "Evaluation of speech impairment in early stages of parkinson's disease: A prospective study with the role of pharmacotherapy," *Journal of Neural Transmission*, vol. 120, no. 2, pp. 319–329, 2013.

[15] P. Di Benedetto, M. Cavazzon, F. Mondolo, G. Rugiu, A. Peratoner, and E. Biasutti, "Voice and choral singing treatment: A new approach for speech and voice disorders in parkinson's disease," *European Journal of Physical and Rehabilitation Medicine*, vol. 45, no. 1, pp. 13–19, 2009.

[16] E. De Angelis, L. F. Mourão, H. B. Ferraz, M. Behlau, P. A. d. L. Pontes, and L. A. F. d. Andrade, "Effect of voice rehabilitation on oral communication of parkinson's disease patients," *Acta Neurologica Scandinavica*, vol. 96, no. 4, pp. 199–205, 1997.

[17] J. Gómez-García, L. Moro-Velázquez, J. D. Arias-Londoño, and J. I. Godino-Llorente, "On the design of automatic voice condition analysis systems. part iii: Review of acoustic modelling strategies," *Biomedical Signal Processing and Control*, vol. 66, p. 102049, 2021.

[18] A. S. Ozbolt, L. Moro-Velazquez, I. Lina, A. A. Butala, and N. Dehak, "Things to consider when automatically detecting parkinson's disease using the phonation of sustained vowels: Analysis of methodological issues," *Applied Sciences*, vol. 12, no. 3, p. 991, 2022.

[19] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.