Identification vocal cord condition during treatment process with dysphonia severity index

To cite this article: M H Trihatmojo et al 2018 J. Phys.: Conf. Ser. 1075 012071

View the article online for updates and enhancements.
Identification vocal cord condition during treatment process with dysphonia severity index

M H Trihatmojo¹, Suprayogi¹ and H Bethaningtyas¹

¹Engineering Physics, Faculty of Electrical Engineering, Universitas Telkom

E-mail: hilmantrihatmojo@gmail.com

Abstract. The purpose of this study was to identify vocal cord condition during treatment process with Dysphonia Severity Index (DSI). The DSI is based on combination of acoustic parameter of MPT(s), F1, Sound Pressure Level (SPL) and jitter that represent vocal quality. Direct measurements of 24 data voice based on 8 patients with vocal cord disorder was taken. The subjects were measured on 3 different day with an interval of 7-9 days. The difference in DSI between each measurement was observed. The mean difference DSI on patients with vocal cord disorder on 3 different day measurements was 0.54, 0.88 and 1.54 respectively. The more positive the patients index, the better vocal quality of the patients.

1. Introduction

The vocal cords are one of the human organs that have an important role in communication process. A person who have a disorder in their vocal cord will have difficulty pronouncing vowel sound clearly in a sustained time [1]. Patients with severe vocal cord disorder will require outpatient care to monitor their health condition or so called monitoring. The purpose of monitoring is to carefully observe a certain condition or behavior so the information obtained could be the basis for making further action decisions. The patient’s medical record information during the outpatient process is required to obtain the results of the physical examination, diagnosis, treatment measures and other services that have been provided to the patient [2]. The invasive method will make the patients uncomfortable to check their health condition [3]. Therefore, a diagnosis based on sound analysis will aid the health worker to detect a disorder of the vocal cords and make the patient more comfortable to monitor their health condition.

Signal speech processing allow the extraction of a set of sound parameters that can be used to diagnose disorder in vocal cords. Currently, the acoustic parameters used in the application of acoustic analysis are the fundamental frequency (F0), jitter, shimmer and HNR[4]. However the measurements of jitter, shimmer, and HNR could not indicate the disorder severity of vocal cord patients [5]. Based on previous research, the research in acoustic analysis of jitter, shimmer, HNR and detection the type of vocal cord disorder using wavelet biorthogonal trasnformation[1] has been done, but the studies to monitor the patient’s cure rate during treatment process with parameter acoustic have not been performed.

Rate of the vocal cord disorder severity could be measured with Dysphonia Severity Index(DSI). DSI is a weighted combination acoustic parameter of Maximum Phonation Time (MPT), Formant 1 (F1), Sound Pressure Level (SPL) and jitter. In DSI index, a positive index value shown a better vocal
quality and negative index value shown the worst vocal quality [5]. The measurements of patients voice quality based on DSI will aid medical personnel to assess the progress of patients health.

2. Methodology

2.1. Human Voice Production Process
The process of sound production in humans can be divided into three physiological processes, i.e. the formation of airflow from the lungs, the change of airflow from the lungs into voiced and unvoiced sound known as phonation and articulation which is the process of modulation or sound settings be a specific sound. Figure 2.1 shows the body organs involved in the production process of human voice [6]. The airflow generated by the lung muscle drive is constant. When the vocal cords are in contraction, the passing airflow makes it vibrate [6]. The airflow will be cut by the movement of the vocal cords into a quasi-periodic pulse signal [7]. Voice signals have characteristics that are formed from a combination of frequency, amplitude and phase. The speaker's voice characteristics have two categories. The first is through the subjective perception of the listener. For example pitch, where one person has a low / heavy base sound while the other is higher or both have the same gender. The second criterion is objective physical characteristics of the sound acoustic emission of sound waves, for example the fundamental frequency with units of Hertz [1].

![Figure 2.1 Human Speech Organs](image)

2.2. Vocal Cord Disorder
The vocal cords are place of vibration in human voice. Uncontrolled open and closure processes in the vocal cords are called vocal cord disorder or dysphonia. The various vocal cords disorder include vocal cord Nodules which are bilateral swell on vocal cords of varying size found on the part with the vocal cord membrane, Laryngopharyngeal Reflux (LPR), a state in which gastric acid moves towards the upper esophagus of the larynx, and Laryngitis which is an inflammation that occurs in the larynx, and others [1].

2.3. Acoustic Parameter
The acoustic parameters used in the application of acoustic analysis are the Maximum Phonation Time (MPT), Formant 1 (F1), Sound Pressure Level (dB), and jitter. The process begins with sound input from a microphone connected to the matlab program and stored in wav format. Furthermore, the calculation of acoustic parameters and DSI values are measured. The calculation of DSI through a combination of acoustic parameters is performed to see the value index. If the value obtained is still below the normal sound standard then the voice in the next week will be re-insert to obtain a comparison of DSI values during the recovery process.
2.3.1. **Maximum Phonation Time (MPT).** Maximum Phonation Time is the maximum time (in seconds) for which a person can sustain a vowel sound when produced on one deep breath at a relatively comfortable sound\(^5\).

![Figure 2.2 Time Domain Signal](image)

Figure 2.2 shows the time domain signal with the x-axis showing the time (s) and the y-axis as Amplitude. The maximum value on the x axis is the MPT value to be measured.

2.3.2. **Formant Frequencies.** Formant frequency is the peak of the spectral sound which is wholly influenced by the articulator. In the processing of speech signals, the formant is an acoustic resonance of the human vocal tract \(^8\). Formant frequencies obtained from the determination at the peak of resonance on the voice signal segment. Formant parameters are determined by "peak picking" on the filter response curve or by solving the square root equation \(^9\). The formant frequency equation can be seen in equation 1, where \(f_S\) is the sample frequency and \(\theta_0\) is the first formant position.

\[
F = \frac{f_S}{2\pi} \theta_0 \text{ Hz} \quad (1)
\]

2.3.3. **Sound Pressure Level.** SPL is a logarithmic measure of the root mean square of sound pressure compared to the reference sound pressure measured in decibels (dB) \(^10\). The sound pressure level equation can be seen in equation 2 where \(P_0\) is the reference pressure in air (\(P_0 = 20 \mu\text{Pa}\)) and \(P\) represents the rms of the measured sound pressure \(^11\).

\[
\text{SPL} = 20 \log\left(\frac{P}{P_0}\right) \text{ Db} \quad (2)
\]

2.3.4. **Jitter.** Jitter represents the difference in the absolute average value between successive periods divided by the average period. Jitter representation can be seen in equation 3 \(^3\). \(T_i\) representing the length of the obtained base frequency period, \(N\) represents the number of basic frequency periods obtained.

\[
\text{Jitter (lokal)} = 1 + \frac{1}{N} \sum_{i=1}^{N-2} |T_i - T_{i+2}| \frac{1}{N} x 100 \% \quad (3)
\]

2.3.5. **Dysphonia Severity Index.** DSI is a multiparameter measurement consisting of a combination of acoustic parameter variables to represent one's vocal quality \(^4\). DSI mathematical equations can be seen in equation 4 with parameters used in DSI is the maximum pronunciation time (MPT) Formant 1, Sound Pressure Level (SPL), and jitter.

\[
\text{DSI} = [0.13 \times \text{MPT} + 0.0053 \times F0\text{(high)} - 0.26 \times I\text{(low)} - 1.18 \times \text{jitter} + 12.4] \quad (4)
\]
3. Result

3.1. Data Collecting
The data were collected by recording the patient's voice at Dustira Cimahi Hospital using Roland Quad-Capture UA-55 soundcard and Shure SM-58 microphone. Direct recording and retrieval of 24 voice data were performed on 8 patients with vocal cords disorder. The subjects were measured three times in a different day, with a time interval of approximately 1 week to monitor the condition of the patient’s vocal cord. Then the DSI value is measured to see the difference on each data measured.

3.2. Data Processing
The data that has been taken is processed using Matlab R2013b to get the value of Maximum Phonation Time (s), Formant 1 (Hz), Sound Pressure Level (dB), jitter, and DSI. The mean values of parameters and DSI on each measurement of control patients can be seen in Table 3.1

| No  | Measurement | MPT(s) | F1 (Hz) | SPL (dB) | Jitter (%) | DSI   |
|-----|-------------|--------|---------|----------|------------|-------|
| 1   | Week 1      | 6.909  | 718.912 | 59.761   | 0.874      | 0.539 |
| 2   | Week 2      | 8.509  | 758.203 | 60.275   | 0.8213     | 0.884 |
| 3   | Week 3      | 12.2   | 822.661 | 60.987   | 0.8037     | 1.541 |

Table 3.1 shows the mean values of the parameters and the DSI value of patients on each measurement. In the 1st week the average DSI is 0.5391, the 2nd week is 0.884 and the 3rd week is 1.5412. The DSI values in the 1st and 3rd measurements show significant differences.

In this study 24 voice data were taken from 8 patients in three different days. In 8 patients with vocal cord disease there are 3 different types of diseases: 1 laryngitis, 5 Laryngopharyngeal Reflux (LPR), 2 thyroids. The development of each acoustic parameter and the DSI value was observed.

3.2.1. MPT Measurement Result. MPT is obtained by taking the maximum phonation time during the recording duration in a sustained voice / a /. The MPT value of each patient during the measurement process had an average value of weeks 1, 2 and 3 of 11,649,12,34,21,38 (s) for laryngitis patients, 6,954, 8,762 and 11.63 (s) for LPR patients as well as 4,425, 5,94 and 9,041 for patients thyroid. For each patient in each measurement there is an increase in MPT value. The increasing trend of value indicates that the sound quality of patients is getting better. MPT view in each patient can be seen in Figure 3.1

3.2.2. Formant 1 Measurement Result. Formant 1 measurements were performed by taking voice data in patients with vocal cord disorder. Formant frequency is an acoustic resonance of human vocal tract [8]. The patient's Formant 1 values had mean values of weeks 1,2 and 3 of 557,352,469,14 and 574,847 (Hz) for laryngitis patients, 753,535, 824,237 and 838,863 (Hz) for LPR patients as well as 713.136, 737.648 and 906,065 (Hz) for thyroid patients.
Figure 3.3 shows the value of Formant 1 in each patient in each measurement. The F1 value of each patient shows the variation in the value of each measurement, indicating that the F1 parameter still has different values for each type of disorder at each measurement.

![Figure 3.2 F1 Parameter (Hz) in Each Measurement](image)

3.2.3. Sound Pressure Level Measurement. The SPL value at each measurement show varying values in each patient with different disorder types. In some patients there was an increasing tendency in SPL values during routine control but in some patients there was an unstable SPL value. The SPL scores for each patient varied with the mean values of measurements at weeks 1, 2 and 3 of 60,496, 57,476, and 58,401 (dB) for laryngitis patients, 60,698, 61,964 and 61,834 (dB) for LPR patients and 57,052, 57,451 and 60,163 (dB) for thyroid patients. Display values of patient SPL parameters during the measurement process can be seen in Figure 3.4.

![Figure 3.3 SPL(dB) Parameter in Each Measurement](image)

3.2.4. Jitter Measurement Result. Jitter measurements are obtained in the same way as other parameters, and the jitter value of patient's during the measurement process was stored. The mean values of the jitter parameters at measurements weeks 1, 2 and 3 were 0.98, 0.941 and 0.92 (%) for laryngitis patients and LPR patients were 0.833, 0.816, 0.767 (%) and 0.924, 0.775 and 0.838 for thyroid patients.

![Figure 3.4 Jitter Parameter in Each Measurement](image)

3.2.5. DSI Measurement Result. DSI measurements were obtained by calculating equation 4 which consisted of combination of Maximum Phonation Time (s), Formant 1 (Hz), Sound Pressure Level (dB) and jitter. The DSI value of each patient showed an increasing trend in each measurement on the different types of disorder with the mean values of weeks 1, 2 and 3 measurements of -0.017, 0.437
and 1.956 for laryngitis patients and in LPR patients of 0.534, 0.834, and 1.376 and DSI of 0.831, 1.232, and 1.747 for thyroid patients.

![Figure 3.5 DSI Parameter in Each Measurement](image)

DSI values can represent the quality of a person's vocal, an increasingly positive DSI score indicating good sound quality [5]. In Figure 3.6 we can see the results of measurements on the DSI values of patients with different types of diseases.

4. Conclusion
Acoustic parameter analysis results show an increase in I-low, F0-High and MPT and decrease in jitter value. With the values of each parameter at weeks 1, 2 and 3 for I-low (SPL) of 59,761, 60,275 and 60,987 (dB), F0-High of 718,912, 758,203 and 822,661 (Hz), MPT of 6,909,8,509 and 12,201 (s) as well as jitter of 0.874, 0.821 and 0.804 (%).

The DSI score of each patient with vocal cord disease shows an increasing trend in each measurement with the lowest DSI value to the highest range (-0.9133 - 2.5212) and the mean values of weeks 1, 2 and 3 are 0.5391, 0.884 and 1.5412.

References
[1] Firda Nurmalida, “Analisa Deteksi Jenis Kelainan Pita Suara Dengan Menggunakan Transformasi Wavelet Biorthogonal”, Tugas Akhir, Universitas Telkom, 2016
[2] Suryadi, Irfani dan Lingga Jaya Bermana Putra. “Sistem Informasi Rekam Medis Rawat Jalan Pada Rumah Sakit Umum Daerah Palembang BARI”.
[3] Achmad, Tri dan Luqman Hakim. “Identifikasi Suara Serak Berbasis Transformasi Wavelet dan ALgoritma Jaringan Syaraf Tiruan” Seminar on Intelligent Technology and its Application, ITS, 2012. Hal : 119
[4] Joao Paulo Teixeira dan Paula Odete Fernandes “Acoustic Analysis of Vocal Dysphonia” Polytechnic Institute of Bragança – Campus Sta. Apolónia 5301 857 Bragança – Portugal,(2015).
[5] Florts L. Wuyts. “The Dysphonia Severity Index: An Objective Measure of Vocal Quality Based on Multiparameter Approach”, Journal of Speech Language and Hearing Research, 2000
[6] Borden, “Speech Science Primer: Physiology, Acoustic, and Perception of Speech”, 1994.
[7] ArryA.A,”ProsesPembentukandanKaracteristikSinyalUcapan”,Teknik Elektro,ITB,Bandung.
[8] Ashish, Bageshree V. Sathe-Pathak. “Extraction of Pitch and Formants and its Analysis to identify 3 different emotional states of a person” Pune University, International Journal of Computer Science Issues, 2012
[9] Fausto, Roy C. Snee, “Formant Location From LPC Analysis Data”, IEEE Transactions On Speech and Audio Processing, 1993.
[10] Errede Steven,”The Human Ear-Hearing, Sound Intensity and Loudness Levels”, UIUC Physics 406 Acoustical Physics of Musics, University of Illinois.
[11] Joseph, Marc Schoenwiesner,”Implementing Loudness Models in Matlab”, Int. Conference on Digital Audio Effects, 2004.