Identification disorders of vocal cords through the public communications VoIP network-based IVR with dysphonia severity index (DSI)

A B Ariyanto¹, F Y Supratman¹ and H Bethaningtyas¹
¹Telkom University, 40257 Bandung, Indonesia

E-mail: ahmad.99bagus@gmail.com

Abstract. The sounds that have been recorded by the system telephone networks Voice over Internet Protocol (VoIP) will be done the noise reduction using the filter process FastICA and Biorthogonal Wavelet. Voice recording process done with a vowel /a/ in one breath for 10 seconds. The result of the process reduction noise sound will be analyzed using the method of Dysphonia Severity Index (DSI) and Harmonic Noise to Ratio (HNR). The DSI method and HNR is used to determine data quality sound healthy and disorder by looking at the value of the both parameter.

1. Introduction
Disorders of the vocal cords causing trouble in talking. The disruption on the phonation organ caused by human vocal cords nodules that are experiencing swelling, irritation or infection [1]. So it takes a Medical Check-Up as early as possible to find out the various possibilities of disease that occur in the vocal cords [2]. Medical check up the vocal cords by doctors in general is using minimally invasive methods, that is to place the laringoskopi and stroboskopi into the mouth of the patient, this would make patients uncomfortable. Limitations equipment in the hospital became the primary factor for patients not to do a check-up. On the other hand, check-up the vocal cords to the hospital requires more time and cost. Based on a lot of problems that arose in the detection of vocal cords through a minimally invasive method, needed another alternative to help the medical world in detecting disease at the vocal cords, one of which is a non-invasive method. This method is done by searching for the value of earmarks in the sound produced by the patient to be a benchmark in determining disease experienced. The technique on the non-invasive method is to ask the patient to pronounce the nodules of vocal/a/ continuously in one breath for a few seconds [3].

2. Methods and Methodology
2.1. VoIP
Voice over Internet Protocol (VoIP) is a technology that is able to transmit the voice data in realtime with the network Internet Protocol (IP) [4]. VoIP technology works by changing the sound analog signal into a digital signal that can be transmitted over a network that utilizes IP. Once converted to digital signals, the noise will be translation into IP packets are then transmitted over a network. VoIP is used in an organization with the aim to reduce the cost of communication and makes it easy to communicate with. In the scale of LAN, VoIP is used to connect one room to another room, then a Metropolitan Area
Network (MAN), VoIP is used to establish lines of communication between the location of one with another location, while a Wide Area Network (WAN), VoIP is typically used to create lines of communication between countries or even continents [5].

2.2. IVR

Interactive Voice Response (IVR) is a computer system that interact with the caller, where the callers will be give the input to the system by pressing the keypad on the phone (DTMF dualtone multifrequency) or by saying a something (a natural language to be known) [6]. The basic principle of IVR system is how the caller read menu and choose an option from the menu to perform an action, or as an alternative to entering information (in the format number via the keypad emphasis).

2.3. FastICA

FastICA is one of the ICA methods. This is allows for a step whitening preliminary for mixed signal zero mean, to increase the speed of convergence of the ICA. Reduction of noise by using FastICA method will give you good results and fast in terms of computation in comparison with other algorithms in a way doing the extraction processed used in Independent Component Analysis [7].

![Figure 1. Mixing of two sources.](image)

Figure 1 shows the mixing of two sources of sound, S is assumed as the source of sound, distance or intensity of the source Matrix mixer is A mixed signal as X, and t is the time. Thus, can be written equations 1 and 2.

\[
\begin{align*}
  x_1(t) &= a_{11}s_1 + a_{12}s_2 \\
  x_2(t) &= a_{21}s_1 + a_{22}s_2
\end{align*}
\]

(1) (2)

In equations 1 and 2, no longer use time dimension of \( t \) in the equation, because ICA components at each observation that presume that they one another ( \( x_1 \) and \( x_2 \) ) are components that are free, as well as the components of the \( s_n \). So the equation can be simplified into the equation 3.

\[
  x_j = a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n
\]

(3)

ICA uses the following equations:

\[
  X = AS
\]

(4)

Where \( S \) is the signal source, parameter A matrix describing the mechanism of incorporation of signal source and signal mixture from the merger describes by \( x \). The basic principle of the ICA is to find the source of the original signal \( S \), and assuming the original signal mutually independent of each other statistically. FastICA iteration scheme is based on the calculation of fixed-point to find the maximum value of WTx nongaussianity [7]. Steps to find a variable value FastICA is as follows:

- Select the value vector of \( w \), can be taken random.
- Find \( w^* = E\{xg(WTx)\} - E\{g'(WTx)\}W \)
- Then locate the variables \( w = w^* \frac{E}{\|w^*\|} \)

with \( W = \) invers matrix \( x \), \( E = \) eigen value, and \( g = \) approximations of negentropy.
2.4. Wavelet Biorthogonal

Wavelet transform is a function generated from one basic function translation and dilatation. Wavelet is used as a tool of analysis of data from a filter, to reduce pollutant factors (noise). Wavelet have many types based on the application, one of which is the Biorthogonal Wavelet [9]. It is Orthogonal Wavelet expansion, which allows for more degrees of freedom. With the addition of the degrees freedom to allow for compose a more symmetric wavelet functions [16]. In the orthogonal wavelet filter analysis and synthesis filter is the inverse of time with one another (for example, \( h(n) = h(\bar{n}) \), \( g(n) = g(\bar{n}) \)). When \( h = h' \) then brings about the relationship between the coeffetien scale and coeffetien wavelet for wavelet orthogonal [10].

3. Vocal Parameter Acoustic

3.1. Dysphonia Severity Index

Dysphonia Severity Index (DSI) was designed as an objective correlation and quantitative of sound quality [11]. It is based on the results of a test run by a Belgium study group analyzing multivariate normal and sound database of pain. Parameters based on the combination of the DSI influence the measurement of the sound basis of the formant frequencies, namely vowel sounds (F1) then low intensity level (in dB), the maximum time the blast (MPT in units of seconds) and Jitter (%). DSI is used to analyze the patient's sound quality for subsequent use as parameters in the determination of the value of votes [12].

\[
DSI = 0.13 \text{MPT} + 0.0053 \text{F1} - 0.26 \text{L} - 1.18 \text{Jitter} + 12.4
\]  

where

- MPT = Aerodynamic Measurements of Variables (s)
- F1 = Formant Frequency Maximum Base (Hz)
- L = Sound Intensity Level (dB)
- Jitter = Time difference from the Two Periods (%)

3.2. HNR

Harmonic to Noise Ratio (HNR) is the acoustic parameters measured as the overall characteristics of the signal, and not as a function of the frequency [13]. The overall value of the signal HNR vary due to the different vocal tract resulting in a different amplitude. The greater the value of the resulting HNR then semkain harmonious voice signal. HNR equation shown in equation (6).

\[
\text{HNR} = 10 \times \log_{10} \frac{A_{Cv}(T)}{A_{Cv}(0) - A_{Cv}(T)}
\]  

4. Result and Discussion

Voice data source used there are two kinds, primary data and secondary data. The primary voice data is the sound that recorded with the procedures directly by patients. While secondary data is the sound that is already available in the database. This secondary data obtained from Computational Physics Laboratory Databases, Telkom University and International Saarbrucken Voice Database (SVD) [14]. The voice data will be recorded through the telephone network to find out the influence of noise caused by communication network. Voice data used in testing as many as 175 final research data, consisting of 115 and 60 healthy data pain.

4.1. The Difference in Healthy and Diseased Sound Characteristics

The sound quality has healthy enough significant differences compared to the sound of misbehaving on the vocal cords. Based on previous research and reference the test results, the value of the parameter of the DSI and the sound of disorders HNR has lower values compared to sound healthy. This is due to the sound of disorders harmony voice signal has a value which is less good compared to sound healthy. Here are the differences parameter DSI and HNR on sound healthy and disorders.
Table 1. Differences of sound the sound of disorders and healthy

| Parameters       | Healthy       | Sick         |
|------------------|---------------|--------------|
|                  | DSI           | HNR          | DSI           | HNR          |
| Average          | 1.169         | 15.804       | 0.139         | 13.789       |
| The Maximum Value| 3.263         | 21.289       | 2.617         | 16.108       |
| The Minimum Value| -0.996        | 10.484       | -4.267        | 11.980       |

Table 1 shows the difference between DSI and HNR sound healthy on the recordings directly. At the sound of the DSI value healthy average of 1.169 HNR value and an average of 15.804. While the sound of disorders DSI value of 0.139 and HNR value average of 13.789.

4.2. The Distinction of Direct Recording and Telephone Network

Sound recording through the telephone network are experiencing a decrease in sound quality due to the presence of noise which causes the addition of foreign signals that may damage the information contained in the data is sound. So, will experience the error detection on sound data.

Table 2. The difference value healthy of DSI and HNR on real-time recording and through the phone

| Parameters          | Directly Before Reduction | Phone Before Reduction | Phone After Reduction |
|---------------------|---------------------------|------------------------|-----------------------|
|                     | DSI           | HNR         | DSI           | HNR         | DSI           | HNR         |
| Average             | 10.253        | 19.001      | 3.935         | 8.522       | 6.707         | 9.513       |
| The Maximum Value   | 16.595        | 27.443      | 8.278         | 16.494      | 12.248        | 18.410      |
| The Minimum Value   | 5.442         | 10.526      | -4.405        | 0.421       | -0.747        | 3.524       |

In table 2 indicates that the sound after recording through the telephone network will result in a decreased in the value. This proves that by doing the recording via the telephone network will provide the noise which caused the value of DSI declined by 61% and HNR of 55.15%. However, after a process of separation noise the sound data back to increase. On the parameters of the DSI increase by 41.3% and the HNR of 10.42%.

Figure 2. Graph Difference HNR and DSI on sound healthy. (a) the HNR. (b) the DSI

Table 3. Difference value of DSI and HNR on real-time recording and telephone at the sound of disorders

| Parameters          | Directly Before Reduction | Phone Before Reduction | Phone After Reduction |
|---------------------|---------------------------|------------------------|-----------------------|
|                     | DSI           | HNR         | DSI           | HNR         | DSI           | HNR         |
| Average             | 7.953         | 13.638      | 4.137         | 2.820       | 6.220         | 4.147       |
| The Maximum Value   | 12.272        | 16.425      | 6.387         | 0.322       | 9.123         | 8.250       |
| The Minimum Value   | -1.136        | 10.278      | -5.018        | 9.531       | -0.588        | 1.583       |
Table 3 shows the difference between the DSI and HNR the sound of disorders. In general the same as healthy sound, there was a decrease after recording is done through the telephone network. The decline on DSI of 47.98%, whereas in HNR amounting to 79.32%. Then the value of the parameter of the DSI and HNR back rise, after reduction of noise. The parameters of the DSI went up by 33.49% and on the parameters of the HNR of 32%. Thus proving that with the reduction of noise data the sound back to the actual information based on recording directly.

![Graph difference of HNR and DSI on sound of disorders. (a) the HNR. (b) the DSI](image)

**Figure 3.** Graph difference of HNR and DSI on sound of disorders. (a) the HNR. (b) the DSI

### 4.3. Analysis Of The Results Of Detection System

The addition of Biorthogonal wavelet filters on the system led to increased sound quality approaching the data at the recording directly. In previous research, has done testing the system noise reduction using FastICA algorithm, difference Table addition of wavelet filter can be seen in Table 4 below [15], [16].

| Table 4 Wavelet Filter Addition Difference |
|--------------------------------------------|
| FastICA | FastICA+Wavelet |
| HNR     | DSI          | HNR    | DSI    |
| 5.114   | 4.728        | 6.285  | 7.690  |

Based on table 4 above, there is an increase in the quality of the existence DSI and HNR addition of Biorthogonal Wavelet filters. Thus, the data will be more sound approach to data recording with sound directly which is the original data without any noise due to communication network.

### 5. Conclusion

Calculation analysis of the data on value DSI and HNR from 175 data through the phone networks recording obtained difference values between the healthy and disorders voice data for system identification. After the recording process via the telephone network, the value of DSI and HNR decline due to noise that causes the change information from the voice signal. DSI value on healthy voice has the range (-4,405-8,278), whereas in HNR is at range (0421-16,494). The disorders sound, data on the value of the DSI sound data lower than healthy. The value of the DSI sound healthy is at the range (5,018-6,387), whereas in HNR is at range (0322-9,531). Then, after the reduction of the value of the DSI on healthy voices are on the range (-0747-12,248) and HNR is at range (3,524-18,410). DSI Value next to the sound of disorders had lower values compared to the sound of disorders, DSI value range to the sound of disorders is (-0588-9,123). While the HNR is at range (1,583-8,250).
6. References

[1] N Yuwono and S Novita 2014 Nodul Pita Suara vol 41 no 6 pp 428–431 (Kalimantan Barat)

[2] Department of Health ,“Attitude and Practice of Medical Checkup,”, Republic of China, 2008.

[3] Arifinto D and Sekartedjo 2004 Speech Disorder Analysis using Time-Varying Autoregressive Proc. IEEE-MWSCAS pp 191-94 (Hiroshima, Japan)

[4] Barry W J and Putzer M 2016 Saarbrucken Voice Database, Institute of Phonitics, Univ. Of Saarland http://www.stimmdatenbank.coli.unisaarland.de/

[5] Wahyuddin M I 2009 Implementasi VOIP Computer to Computer Berbasis Freeware Menggunakan Session Initiation Protocol Jurnal Artificial (ICT Research Center UNAS)

[6] M Imam 2015 Aplikasi Reduksi Noise dalam Perbaikan Kualitas Suara untuk Deteksi Gangguan Pita Suara (Universitas Telkom)

[7] H Aapo and E Oja 2000 Independet Component Analysis: Algorithm and Application (Helsinki University of Technoligy)

[8] A Niknejad 2005 Noise in Analog Comunication System (UC Barkeley: University of California)

[9] N Firda 2016 Analisa Deteksi Jenis Kelainan Pita Suara dengan Menggunakan Transformasi Wavelet Biorthogonal (Universitas Telkom)

[10] Keinert Fritz 1996 Biorthogonal Wavelets For Fast Matrix Computations I (owa State University)

[11] M S De Bodt and F L Wuysts. 2010 The Dysphonia Severity Index (DSI) (University of Antwerp)

[12] M. Farrús, J. Hernando, and P. Ejarque 2007 Jitter and Shimmer Measurements for Speaker Recognition (Spain: Technical University of Catalonia)

[13] T. Wahyu and R. Ningsih 2015 Nilai Forman dan Karakteristik Bunyi Ujaran Pada Autism Spektrum Disorders (ASD) dam Anak Normal vol 6 no 1972 pp 125–133 (Universitas Gunadarma)

[14] Mifflin 2002 American Heritage Dictionary of the English Language, 4th Edition (Houghton Mifflin Company)

[15] A. Cohen, Daubechies I and Feauveu 1990 Biorthogonal Bases of Compactly Supported Wavelets (Universite Paris)

[16] Wuysts F L, Geert M and Marc R 2010 The Dysphonia Severity Index: An Objective Measure of Vocal Quality Based on a Multiparameter Approach (Belgium: University of Antwerp)