Spectral Analysis of Familiar Human Voice Based On Hilbert-Huang Transform

Agustinus Bimo Gumelar
Fakultas Teknologi Elektro
Institut Teknologi Sepuluh Nopember
Fakultas Ilmu Komputer
Universitas Narotama
Surabaya, Indonesia
bimogumelar@narotama.ac.id /
bimogumelar@ieee.org

Mauridhi Hery Purnomo
Fakultas Teknologi Elektro
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
hery@ee.its.ac.id

Eko Mulyanto Yuniarno
Fakultas Teknologi Elektro
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ekomulyanto@ee.its.ac.id

Indar Sugiarto
Fakultas Teknik Elektro
Universitas Kristen Petra
Surabaya, Indonesia
indi@petra.ac.id

***Abstract*— Spectral analysis of human voice signals is important to reveal hidden information when is not available in the time-domain. Extracting spectral information from those voice signals will enhance our knowledge in understanding the nature and characteristic of the voice. It concerned with the decomposition method of voice signals into simpler components in frequency and time. The frequency analysis tools are also give beneficial for describing the spectral distribution in a voice signal, very often the methods used by the tools have limitations that restrict us to interpret the data properly. This paper describes a powerful data analysis method called the Hilbert-Huang transform (HHT), which can be used to extract audio frequency components from nonlinear and nonstationary human voice signals. It can describe the audio frequency components locally and adaptively for nearly any oscillating signal. This makes it very extremely versatile to be used for analysing familiar human voices.**

Keywords—Spectral Analysis, Human Voice Analysis, Hilbert Huang Transform, Hilbert Spectrum

I. Introduction

Human voice signals have inherently complicated nonstationary and non-linear waveforms. Human voice is the most important mean to pass information from one person to another. Passing messages by voice is the most effective way for human communication. In this digital information era, people can doing voice messaging more effectively, which leads to the promotion of social development. In this paper, we present our method for analysing voice signals. It is based on a voice collection system, which collects human voice signal and analyzes the underlying information from the spectrum of the signal. Extracting the spectral information from the signal enhances our knowledge and understanding that is important for further analysis to uncover the hidden information from the signal. Two common methods for analysis of non-stationary signals are Wavelet Transform and Fourier Transform[1]. However, both methods have serious limitation when applied to voice signals. We propose to use a Hilbert-Huang Transform (HHT) as an alternative method to analyse human voice signals.

In order to understanding human voice, basicly is an embodiment of human itself in many social aspects [2]. Human voice pattern are a significant part of our identity. A person’s voice is one of the most fundamental attributes that enables communication within others in physical environment, or at remote locations by using mobile phones or teleradio systems, also it can be using from digital media. Familiar human voice actually so easy to understand, even in some feature like the resonance was manipulated, it still easier to understand when the context of conversation or the words spoken by a stranger. When we talk with other people using voice, it subsequently we become familiar with their voices, and by this enables us to recognize those people by their voice[3]. The familiarity and voice representation based on perceptual representation become unique and and give a signal used for recognition of individual speakers [4]. However, without the knowledge of them, people often leave traces of their personal vocal identity in many different time, situation, either in scenarios and contexts.

This paper is organized as follows. Section I discusses on how important of human voice pattern. On Section II, which is briefly present of review from previous methods of different time-frequency analysis, and then discuss on Hilbert-Huang Transform to develop spectral analysis which is followed in Section III. Some discussion about interesting results that derive from the experimental simulation. This topic presenting the different distance and similarity measures used in this research, which are discusses in Section IV. At last in the Section V provides some final conclusions and directions for future work..

II. Time-frequency Analysis Methods

There are many methods available for processing stationary and non-stationary data, and most of them actually depend of Fourier based analysis. There are also limited only to solve on linear system. In this section, briefly, we will review for non-stationary data processing methods as follows.

A. Fourier Transform

Fourier analysis is become commonly member of mathematical function and techniques. From the decomposing signals into sinusoids function are become its techniques.

978-1-5386-7509-0/18/$31.00 ©2018 IEEE

Fourier analysis has been become dominant and principal analytical tool in voice signal processing. From fundamental nature of sinusoids, and the perspective of periodicity, we could recognize that Fourier analysis as a theory whereby quite general phenomena, whether or not they are themselves periodic or exhibit any obvious overall cyclical behavior, can be understood as combination of basic periodic elements of definite frequencies [5]. In particular Fourier analysis, gives many research contribution to investigate, identify, and techniques that are useful in signal analysis. And this contribution has a large number of practical applications. In order to derive an equation from Fourier formula, as it follow:

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi f t} dt \qquad (1)$$

Since the Fourier transform of time varying signal is a time independent function, it does not register frequencies varying with time. For spectral information to be useful, it should be possible to identify regions in time corresponding to the desired spectral characteristics of a time varying signal.

B. *Wavelet Transform*

Nowadays, wavelet method has wide range of use in the present scientific computation. In many aspect of computational process, different types of tasks could be done perfectly by using wavelet method, for example in biometric recognition. Voice signal, especially human voice is a highly tremendous non-stationary signal. Before then, the Fourier Transform could not be a useful tool for special purpose analysis. Wavelet Transform have a unique approach for analysing nonstationary signals, as it is useful in localizing pattern and give an indication both in time and frequency scales. Wavelet transform become a state-of-the-art's tool that could cuts up data or functions into different frequency components, and then from each component came with a resolution matches to its scale (wavelet). Wavelets also could give a better signal representation by using multiresolution analysis, both in time and frequency domain. By following the general definition, Wavelet Transform are an adjustable window of Fourier spectral analysis. The Wavelet Transform could be defined by the formula as it follows [6]:

$$\tilde{s}_{\psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt, \qquad (2)$$

The formula shows that Wavelet spectrum ψ could carries time and frequency and being an important function. By using two arguments function $\tilde{s}_{\psi}(a,b)$, could represents a spectrum $s(t)$. From scale showed on a parameter and it correlates with a frequency. The b parameter is using for time based translation.

C. *Wigner-Ville Distribution (WVD)*

In the physical condition, it is common to get a periodic prediction or to get some preventive state. This condition usually generates a signature from instant or delaying time in non-stationary mode. So, a time-frequency based on domain representation would be needed to characterize such as those signatures. A method to represent the time-dependent events in non-stationary modes. By following Gabor's function, the Wigner-Ville distribution could recognize the insufficiency of time and frequency analysis. The Wigner-Ville distribution[7] has indicated that a signal has a spectral structure, and they called an instantaneous spectrum. That spectrum had a physical attribute from energy density.

$$W(t,f) = \int_{-\infty}^{\infty} z\left(t + \frac{\tau}{2}\right) . z^{*}\left(t - \frac{\tau}{2}\right) e^{-j2\pi f \tau} d\tau \qquad (3)$$

When the complex signal corresponds to real signal, there would be taking from the Hilbert Transform of the real signal. In consequence, the WVD was done with quantum mechanism, by the real joint distribution of the signal, both in time and frequency domain. It could achieve optimal energy in the time-frequency based.

III. Hilbert-Huang Transform (HHT)

As the Hilbert-Huang Transform (HHT) method is still in developing stage. There are many scientific application such as image processing [8], biomedical signal interpretation [9], fault diagnosis technique [10] using HHT method. In this paper we will introduce the Hilbert-Huang Transformation as a new technique which can be applied to human voice signal decomposition into two components. The first component is empirical mode decomposition (EMD) method and second is Hilbert spectral analysis.

A. *Empirical Mode Decomposition (EMD)*

In audio or voice analysis using spectral analysis, there are many methods have been introduced and developed. Mostly, there are based on an assumption of linearity and stationarity of the signal data. Therefore, this open condition gives a new challenge to analyze the voice signal data to deal with non-linear and non-stationary data. Furthermore, to capture the characteristics as the function of the time-domain, many researchers develop time-frequency based analysis, which is generated a spectrum at the discrete time. The empirical mode decomposition (EMD) method is developed by Huang et al. [11], for each signal on EMD $x(t)$could be decomposed by:

$$x(t) = \sum_{j=1}^{n} c_j + r_n \qquad (4)$$

The following c_j is an Intrinsic Mode Functions, which is when one iteration from IMF could achieve a decomposition value of the voice signal into n value of empirical modes, it also gives a residue value r_n. The most important that will be the characteristics of the EMD is to decompose the multi-component voice signal into a number of single component voice signals. It decomposes the voice signal into the sum of several IMF integrity and orthogonality [12], also decomposes according to the signal itself and the number of the IMF when it is finite.

B. *Intrinsic Mode Functions (IMFs)*

In audio frequency has a very important feature, called the instantaneous frequency. This feature used for the non-stationary signal. The instantaneous frequency is the rate of change of the phase $\theta'(t)$ [11].

Physically, the necessary conditions for us to define a meaningful instantaneous frequency are that the functions are symmetric with respect to the local zero mean, and have the same numbers of zero crossings and extrema. Based on these observations, Huang et al. [13] defined IMF as a class of functions that satisfy two conditions: 1) In the whole voice signal data set, the number of extrema and zero-crossings must be equal or differ at most by the others. 2) At any point in the signal data, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

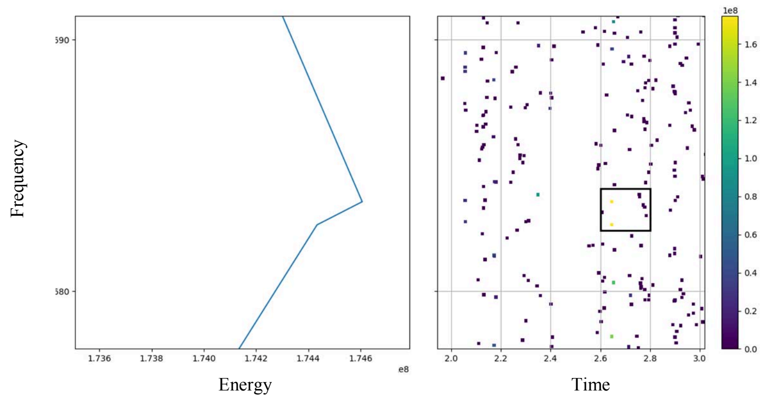


Fig. 1. Square of energy (amplitude) on Hilbert Spectrum ($S\_(i,j)$) Scatterplot

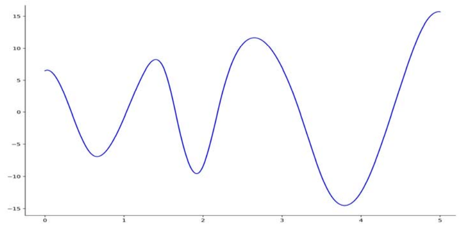


Fig. 2. Voice signal result from IMF process in an oscillatory modes

By Intrinsic Mode Function (IMF) could satisfies the condition at any time are instant, the mean value of the upper envelope as defined by the local maxima and the lower envelope as defined by the local minima is zero. The algorithm for IMF, shown below:

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + r_n(t) \tag{5}$$

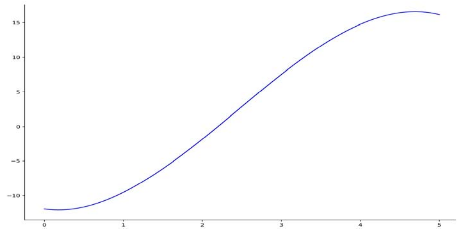


Fig. 3. Result of residue value from IMF

The process on Fig. 2. looping by outer iteration until the residual reaches with sufficiently small value. Residual value represents the portion of the original signal X were not decomposed by EMD. The end of decomposition process is residue value $r_n(t)$.

C. Hilbert Spectrum

In Hilbert Spectrum, the energy production from energy density distribution in a time-frequency space divided into equal size bins of $\Delta t x \Delta\omega$ with the value in each bin appointed as $a^2(t)$ at the proper time (t) and proper instantaneous frequency (ω) [14].

Currently, the energy spectrum from Hilbert Spectra is not defined in terms of density. The value is simply the energy value at particular bin. Therefore the value in each bin should be defined as follow:

$$a\_(i,j)/(\Delta t\,.\,\Delta\omega) \text{ for amplitude spectra and} \tag{6}$$

$$(a\_(i,j)\hat{}2)/(\Delta t\,.\,\Delta\omega) \text{ for energy spectra} \tag{7}$$

To compute the Hilbert Spectrum in the time-frequency space, two variable for each of them designated as follows:

$$time = t_0, t_0 + \Delta t, t_0 + 2\Delta t, \ldots t_0 + i\Delta t, \ldots t_1 \tag{8}$$

$$freq = \omega_0, \omega_0 + \Delta\omega, \ldots, \omega + j\Delta\omega, \ldots, \omega_1 \tag{9}$$

The values and sizes of time and frequency variables could be selected to fit by given restrictions on t_0 *and* t_1 have to reside within the interval of voice signal data span $[0, T]$ and Δt value cannot be smaller than the sampling step [14]. The sampling rate limits the frequency scale on the Fourier Spectrum, but in Hilbert spectral analysis, the frequency values are continuous and have no limitation to its range of values. By designating the values at $t_i = t_0 + i\Delta t,\ \ \omega_j = \omega_0 + j\Delta\omega$ as the value to $S_{i,j}$ as shown on Fig. 1. The Hilbert spectrum could be defined as follows:

$$Hilbert(\omega_j, t_i) = S_{i,j}$$

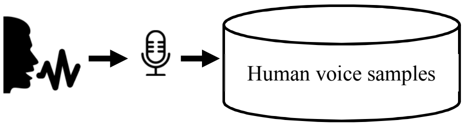
$$S_{i,j} = \frac{1}{\Delta t\,.\Delta\omega} \sum_j \sum_i a_{i,j}^2 \tag{10}$$

IV. Experimental Simulation

Now, we present an experimental work that we conduct to obtain hidden information from human voice signals in the time-domain by extracting based on Hilbert-Huang Transform.

A. Familiar Human Voice Dataset

To build the dataset for our studies, we collected a voice audio samples from a group of call center agent speakers. There were around 20 samples for both the male and the female call center agents that took each user approximately half an hour to record. Among these, we picked ten male and ten female voice, in which all clips are successfully recorded in a non-noisy environment and created with a similar style and pace of the original speakers.



$\boldsymbol{Voice = (p\_1 \ldots p\_n)}$

Fig. 4. Recording process to collect audio voice samples

In this paper, we conduct voice recording for the first phase involves 20 collection of voice samples both of male and female $Voice = (male_1, female_1 \ldots male_{20}, female_{20})$ The original voice data were used to input human voice signal consisting of a sum of sine waves sampled at 44.1 kHz. The human voice signal in a file represents a series of samples that capture the amplitude of the sound over the time frame and storing in waveform file. The second phase is resampling the voice audio sample at the new sample rate.

B. HHT Process

The HHT process begins by establishing the IMF with the EMD (Empirical Mode Decomposition) method from the input signal shown in Fig. 4. After obtaining the IMF, HHT process was carried out for each IMF, the process which shown in Table 1 in line 14-18. Signal analysis is used to analyze that there will be several variables, including phases, instant frequencies, and amplitude. The process to get phases are shown in Table 1 in line 20-26, then to get the frequencies shown in Table 1 in line 27-31, and to get phases shown in Table 1 in line 32-38. After that, the end of the process by making Hilbert spectrum plot by using time variables, instant frequencies, and the resulting amplitude values.

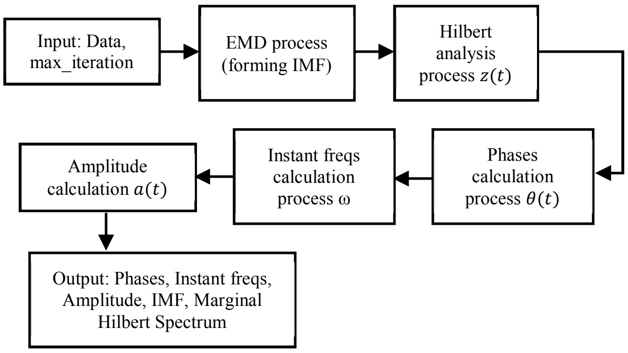

Input: Data, max_iteration
EMD process (forming IMF)
Hilbert analysis process $z(t)$
Phases calculation process $\theta(t)$
Instant freqs calculation process ω
Amplitude calculation $a(t)$
Output: Phases, Instant freqs, Amplitude, IMF, Marginal Hilbert Spectrum

Fig. 5. HHT Process

TABLE I. HILBERT-HUANG TRANSFORM PSEUDOCODE

HHT pseudocode

Input: Voice file (wav extension)
Output: Hilbert spectrum image
IMF: List IMF and final residue
frequencies: list instant Frequency
amplitudes: list instant Amplitude
phases: list instant Phase
Hilbert_signal: list The Hilbert transformed signal(contain real and imaginer, $x + iy$)
x: real value of Hilbert_signal[i]
y: imaginer value of Hilbert_signal[i]
i: index
t: list time
dt: delta time (t_n-$t_{(n-1)}$)
temp_t: container for a while
initial state: temp_t=0

```
1   get numeric value of voice file, channels, duration and sample
2   rate: signal=numeric value voice file
3   if channels ==1 then
4           IMF = Process decompose IMF
5   Else
6           Choose 1 channel
7           IMF = Process decompose IMF
8   End if
9   For i = 0 to i < length of signal list do
10          temp_t = temp_t + (duration/ length of signal)
11          t[i]=temp_t
12  End for
13  or i = 0 to i < length of IMF list do
14  For j = 0 to j<length of t do
15  Hilbert_signal[i][j] = IMF[i][j] + I$\frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{IMF[i][\tau]}{t-\tau} d\tau$
16          End for
17  End for
18  For i = 0 to i < length of Hilbert_signal list do
19          For j = 0 to j<length of hilbert_signal[i] do
20                  x= real value of hilbert_signal[i][j]
21                  y= imaginer value of hilbert_signal[i][j]
22                  phases[i][j]= arctangent of (x/y)
23          End for
24  End for
25  For i = 0 to i < length of phases list do
26  For j = 0 to j<length of phases [i] - 1 do
27          Frequencies[i][j]=(phases[i][j+1]- phases[i][j+1])/2*π*dt
28          End for
29  End for
30  For i = 0 to i < length of Hilbert_signal list do
31          For j = 0 to j<length of hilbert_signal[i] do
32                  x= real value of hilbert_signal[i][j]
33                  y= imaginer value of hilbert_signal[i][j]
34                  amplitudes[i][j]= $\sqrt{(x^2+y^2)}$
35          End for
36  End for
37  Plot hilbert spectrum(t, frequencies, amplitudes)
38
```

C. Empirical Mode Decomposition Method

Properties of EMD Basis were adaptive basis based on and derived from the data by empirical method to satisfy nearly all the traditional requirements for basis a posteriori.

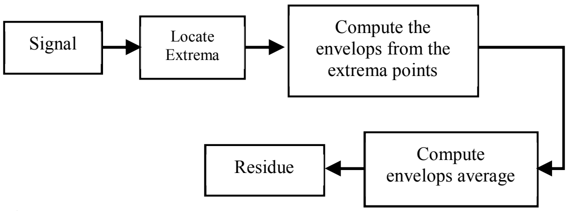


Fig. 6. EMD Process

TABLE II. DECOMPOSE PSEUDOCODE

Decompose pseudocode

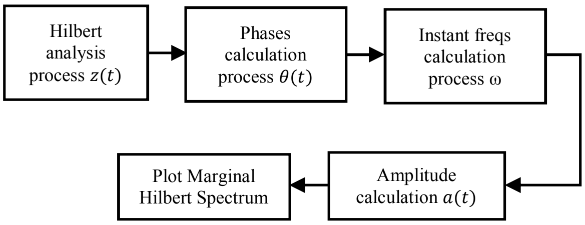
input: signal,max_iterasi
output: all available IMF and residue
IMF: list IMF and final residue
Ext_res: local extrema
n= iteration
Initial state: residue = signal, n=0

```
While stoping condition not satisfied do
        residue = signal-IMF
        while true do
                find local extrema, ext_res = local extrema
                    If n >=max_iterasi then
                    break
                create upper and lower envelops from ext_res
                computer envelops average(upper and lower),
                    avg=upper-lower/2
        IMF = IMF - avg
        end while
            residue = residue - IMF
                if final residue then
            break
        end if
end while
```

D. Hilbert Spectral Analysis

The overall process of HHT is to present a spectral analysis in time series data for providing the description of time-frequency domain. The HHT method also describes non-stationary data locally [15]; it shows in Equation 11.

$$H[x(t)] \equiv \hat{y}(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \qquad (11)$$



Hilbert analysis process $z(t)$
Phases calculation process $\theta(t)$
Instant freqs calculation process ω
Amplitude calculation $a(t)$
Plot Marginal Hilbert Spectrum

Fig. 7. Hilbert Analysis Process

E. Distances and Similarities Measurement

To compare data points, we can quantify how similar the data with similarity or affinity metric, or we can quantify how different the data with a dissimilarity or a distance metric. There are many possible metrics methods like Mahalanobis, Hamming, Gaussian, Jaccard. It is sometimes useful to consider several different metrics and then combine the methods. To get the value from IMF and the real signal, six different methods were used: RMSE, MAPE, Correlation, Manhattan, Cosine, and Euclidean.

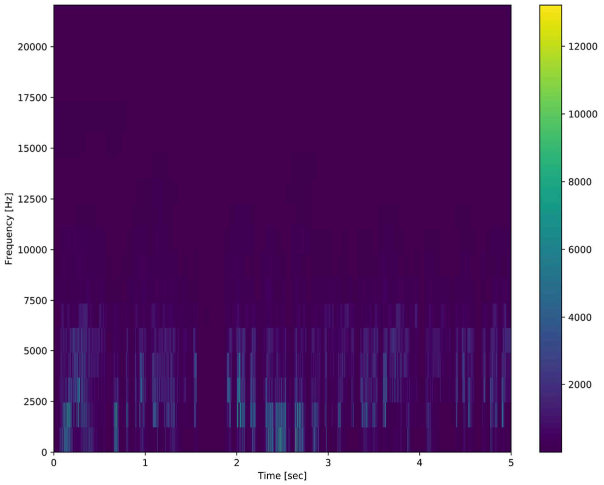


Fig. 8. Plot of Hilbert Spectrum

1) RMSE is a prediction model to estimate variable [16] *X_{model}* is defined as the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{obsj}-X_{modelj}\right)^{2}}{n}} \qquad (12)$$

X_{obs} is observed values and X_{model} is modeled values at time/place *i*. RMSE become frequently used to measure of the difference between values predicted by a model and the values actually observed from the environment that is being modeled [17]. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power.

2) MAPE is time series based function to forecast homogeneous data [18], [19]. When there are no extremes to the data, MAPE functions would be the most common measure of forecast error. The function will measures the accuracy for constructing fitted time series values in statistics and the time series must be identical in size. The actual and estimated time series defined as follows :

$$MAPE = \frac{100}{N} . \sum_{i=1}^{N} \left|\frac{x_i-\widehat{x_i}}{x_i}\right| \qquad (13)$$

3) Euclidean Distance is the most simple and commonly used distance between two points in the multidimensional space and is an extension to Pythagoras theorem[20], [21]. In simple terms, the Euclidean distance is defined as:

$$Euc\ dist = \sqrt{\sum_{k=1}^{n}(p_k-q_k)^2} \qquad (14)$$

4) Cosine Similarity is generally used as a metric for measuring distance and determined by cosine function [22]. Similar are linearly correlated between the voice signal. Cosine similarity is utilized to measure the similarity between pairwise audio vectors in the middle [13]. Given the cosine similarity represented as follows :

$$\cos similarity = \frac{X_1.X_2}{\|X_1\|\|X_2\|} = \frac{\sum_{i=1}^{n} X_1 X_2}{\sqrt{\sum_{i=1}^{n} X_1^2 . \sqrt{\sum_{i=1}^{n} X_2^2}}} \qquad (15)$$

5) Manhattan Distance can be defined as distance between two points in Euclidean space with fixed Cartesian coordinate system [23]. It is the sum of the lengths of the projections of the segment between the points into the coordinate axes. In simple terms, it is defined as [24] :

$$d(i,j) = \sum_{k=1}^{n} |x_k(i) - x_k(j)| \qquad (16)$$

6) Correlation, is one of calculation metric based on statistical measure technique that describe the strength of association between random variable. The correlation coefficient (r) is cosine $\propto$ where the origin of the coordinate system is the mean species composition of a sample unit in the centroid [25].

$$r\_distance = (1-r)/2 \qquad (17)$$

TABLE III. DISTANCE AND SIMILARITIES RESULT METRIC

IMF	RMSE	MAPE	Corr	Manhat	Cosine	Euclid
1	1885.027	160.21	0.516	244727515.9	-30.540	885160.2
2	1633.917	190.05	0.347	216445223.4	-41.577	767245.5
3	1952.588	224.32	0.578	246999546.8	-26.50	916885.1
4	2051.928	234.12	0.694	259434365.3	-18.920	963532.7
5	2081.771	241.15	0.742	257648036.7	-15.771	977545.9
6	2142.76	173.11	0.893	264356783.6	-5.974	1006185
7	2155.415	126.37	0.993	268165342	0.598	1012127
8	2154.451	116.56	0.999	268145879.4	0.963	1011675
9	2154.183	113.48	0.999	268053453.6	0.985	1011549
10	2153.701	108.26	1.085	267887050.5	1.059	1011323
11	2153.572	108.45	0.928	267918201.8	0.990	1011262
12	2153.529	114.84	0.998	267966793.9	0.965	1011242
13	2153.562	125.03	0.997	268082710.7	1.008	1011257
14	2153.588	125.84	0.996	267979014	0.9961	1011270
15	2153.425	110.54	1.025	267860344.9	1.0094	1011193
16	2153.429	106.05	1.029	267812720.2	1.035	1011195
17	2153.425	103.56	1.028	267753518.3	0.955	1011193

As seen in the Table III, it shows the smallest amount of distance value in every method represent better similarities or merely close on distance between processed signal and real signal.

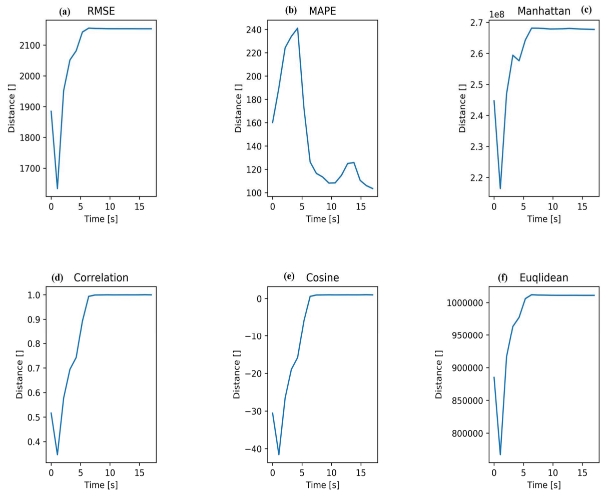


Fig. 9. Distance and similarity result from IMF value on clockwise: (a) RMSE; (b) MAPE; (c) Manhattan; (d) Correlation; (e) Cosine; (f) Euclidean

V. Conclusion

This paper propose to extract human voice signal features based on Hilbert-Huang Transform (HHT). With HHT we can analyze the time-frequency domain and the voice signal data can be decomposed into several function that give a meaningful remark from instantaneous frequencies. From the Hilbert Spectrum we can observe the frequency characteristics of the experimental data. Several distance metrics have been considered for measuring spectral data from intrinsic mode function of human voice data database, and then compared. Every value give insight to determine the distance and similarity between original signal and the processed signal. With EMD, any complicated signal can be decomposed into a finite number of simple signals, each of which includes only one oscillatory mode in any time location. To confirm the obtained results and the effectiveness of the approach, pre-processing strategy must be evaluated with a large dataset of real signals and in different experimental situation and conditions. Though different methods were proposed in this experiment, a better approach with an efficient distance metric that gives good clustering results and will be runs faster in future research work.

References

[1] C. Lin, H. Chuang, Y. Wang, and C. Jian, “HHT-Based Time-Frequency Analysis in Voice Rehabilitation,” *Biotechnology*, pp. 49–53.

[2] D. Sidtis and J. Kreiman, “NIH Public Access,” *Integr. Psychol. Behav. Sci.*, vol. 46, no. 2, pp. 146–159, 2012.

[3] R. Port, “How are words stored in memory? Beyond phones and phonemes,” *New Ideas Psychol.*, vol. 25, no. 2, pp. 145–172, 2007.

[4] A. Andics, J. M. McQueen, K. M. Petersson, V. Gál, G. Rudas, and Z. Vidnyánszky, “Neural mechanisms for voice recognition,” *Neuroimage*, vol. 52, no. 4, pp. 1528–1540, 2010.

[5] “Part I Fourier analysis and applications to sound processing,” .

[6] M. Ziółko, R. Samborski, J. Gałka, and B. Ziółko, “Wavelet-Fourier Analysis for Speaker Recognition,” no. September, pp. 1–6, 2011.

[7] by T. A C M Claasen and W. F. G Mecklenbräuker, “THE WIGNER DISTRIBUTION-A TOOL FOR TIME-FREQUENCY SIGNAL ANALYSIS PART I: CONTINUOUS-TIME SIGNALS,” 1980.

[8] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, “Image analysis by bidimensional empirical mode decomposition,” *Image Vis. Comput.*, vol. 21, no. 12, pp. 1019–1026, 2003.

[9] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung, “Use of intrinsic modes in biology: examples of indicial response of pulmonary blood pressure to +/- step hypoxia.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 22, pp. 12766–71, Oct. 1998.

[10] S. J. Loutridis, “Damage detection in gear systems using empirical mode decomposition,” *Eng. Struct.*, vol. 26, no. 12, pp. 1833–1841, 2004.

[11] S. Tolwinski, “The Hilbert Transform and Empirical Mode Decomposition as Tools for Data Analysis Real Signals and the Hilbert Transform,” *Transform*, no. 1, pp. 1–18, 2007.

[12] I. Magrin-Chagnolleau and R. G. Baraniuk, “Empirical mode decomposition based frequency attributes,” *Proc. 69th Soc. Explor. Geophys. Meet.*, pp. 1949–1952, 1999.

[13] H. Li, Y. Zhang, and H. Zheng, “Hilbert-Huang transform and marginal spectrum for detection and diagnosis of localized defects in roller bearings,” *J. Mech. Sci. Technol.*, vol. 23, no. 2, pp. 291–301, 2009.

[14] N. E. HUANG, X. CHEN, M.-T. LO, and Z. WU, “on Hilbert Spectral Representation: a True Time-Frequency Representation for Nonlinear and Nonstationary Data,” *Adv. Adapt. Data Anal.*, vol. 03, no. 01n02, pp. 63–93, 2011.

[15] N. E. Huang *et al.*, “On holo-hilbert spectral analysis: A full informational spectral representation for nonlinear and non-stationary data,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016.

[16] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature,” *Geosci. Model Dev*, vol. 7, pp. 1247–1250, 2014.

[17] S. Beheshti, M. Hashemi, E. Sejdic, and T. Chau, “Mean Square Error Estimation in Thresholding,” *IEEE Signal Process. Lett.*, vol. 18, no. 2, pp. 103–106, Feb. 2011.

[18] R. Adhikari and R. K. Agrawal, “A Homogeneous Ensemble of Artificial Neural Networks for Time Series Forecasting,” 2011.

[19] J. D. Hamilton, “Time Series Analysis,” *Book*, vol. 39, no. 1. p. xiv, 1994.

[20] M. Greenacre and R. Primicerio, “Measures of distance between samples: Euclidean,” *Multivar. Anal. Ecol. Data*, pp. 47–59, 2013.

[21] E. San Segundo, A. Tsanas, and P. Gómez-Vilda, “Euclidean Distances as measures of speaker similarity including identical twin pairs: A forensic investigation using source and filter voice characteristics,” *Forensic Sci. Int.*, vol. 270, pp. 25–38, 2017.

[22] A. Huang, “Similarity measures for text document clustering,” *Proc. Sixth New Zeal.*, no. April, pp. 49–56, 2008.

[23] M. Helén and T. Virtanen, “Audio query by example using similarity measures between probability density functions of features,” *Eurasip J. Audio, Speech, Music Process.*, vol. 2010, 2010.

[24] N. Paltz, “QUERY-BY-EXAMPLE Tuomas Virtanen and Marko Hel ́ en Tampere University of Technology,” pp. 82–85, 2007.

[25] B. McCune, J. B. Grace, and D. L. Urban, “Distance Measures,” *Anal. Ecol. Communities*, p. 304, 2002.