Analysis Discrete Hartley Transform for the recognition of female voice based on voice register in singing techniques

Muhathir¹, R Muliono¹, N Khairina¹, M K Harahap², S M Putri³
¹Department of Informatics Engineering, Universitas Medan Area, Medan, Indonesia
²Department of Informatics Engineering, Politeknik Ganesha, Medan, Indonesia
³Department of Electrical Engineering, Universitas Medan Area, Medan, Indonesia

E-mail : muhathirbangdes@gmail.com, rizkimuliono@gmail.com, nurulkhairina27@gmail.com, choir.harahap@yahoo.com, syarifahmuthiaputri@gmail.com

Abstract. Automatic speech recognition is the ability to receive and identify spoken words by converting analog signals to digital and extracting unique vocal characteristics such as pitch, frequency, tone, and rhythm to form speaker models or sound samples. The voice sample used is the voice register, the voice register is the division of the area of the human voice based on the source of the sound, the sensation of the resonant space, shape, color, sound timbre, and the high and low tone produced. Discrete Hartley Transform is used as a transformation to process the sound sample to be classified. DHT, DHT + High Pass Filter and DHT + Low Pass Filter in transforming voice register signals can only classify with an average true positive rate of 69.67%. The addition of the filter does not affect the classification results because the sound frequency used is in ideal conditions so that there is no noise that affects the classification results of the voice register.

Keywords : DHT, Voice Register, Low Pass Filter, High Pass Filter.

1. Introduction

The sound is the most effective means of communication between humans. Besides being effective, humans are also more familiar with using voice in communication. Sound processing is a very important concept for all types of systems that require human interaction in daily activities [1]. Sound processing is the ability to receive and identify spoken words [2]. Automatic speech recognition converts speech signals into word sequences with the help of algorithms implemented in computer programs [3]. The aspects of human voice behavior are used for identification by changing spoken phrases from analog to digital format, and extracting unique vocal characteristics such as pitch, frequency, tone, and rhythm to form speaker models or sound samples [4].

According to [1] the sound processing system can be categorized into four types: Isolated Words, Connected Words, Continuous Speech, Spontaneous Speech. Isolated Words: is the processing of sound by receiving one word at a time where the system hears the speaker to wait between sayings, Connected Words: the word system is connected almost the same as Isolated Words, but allows separate utterances to be the same turn with minimal pauses between utterances, Continuous Speech: a continuous sound processing allows the user to speak almost naturally, while the computer determines the content (basically this computer dictation). Spontaneous Speech: is a spontaneous sound that can be categorized
as an audible and untrained natural sound. A spontaneous sound processing system must be able to handle various natural sound features such as the words "ums" and "ahs" and even a little stutter. Sound processing usually only classifies or identifies the owner of the voice, in this paper will be analyzed for women's voice registers in singing songs using Discrete Hartley Transform (DHT), in singing techniques usually women will use Chest Voice, Head Voice, Falsetto, and Vocal fry voice registers[6].

2. Review of literature

2.1 Voice Classification
Various methods are used in classifying sounds, such as Artificial Neural Networks (ANN), DCT, Gabor Filters, Statistics, Wavelets, and others. From human speech can be obtained various information including: Speech Identity, Expression, Dialect / Slog / Tribe, Gender Type, Distance of sound sources, Voice rate, Age, Word Recognition, Loud level, Saturation rate, Health condition and language quality[6].

2.2 Voice Register
The voice register is the division of the area of the human voice based on the source of the sound, the sensation of the resonant space, shape, colour, sound timbre, and the resulting low and low pitch [7] [8]. Some voice registers:
   a. Vocal fry
      Vocal Fry is a human voice produced through the vibrations of the glottis or vocal cords where the two vocal cords blend together perfectly but are weak with a slight boost of air from the lungs.
   b. Chest Voice
      Chest voice is the sound produced when the resonance chamber occurs in the oral cavity or experts call it in the chest.
   c. Falsetto
      Falsetto is the head voice whose resonance does not reach the nasal cavity or the head. The resulting sound is heavy, somewhat nasal like a cork, and the resulting tone is not as high and not as sound as the head voice.
   d. Head Voice
      The head voice is the sound produced when the resonance chamber occurs in the nasal cavity or head. The head voice character itself is light, soft, loud, crunchy, more melodious than falsetto, and the tone produced is also higher than falsetto.

2.3 Pre-Emphasize
Pre-emphasis is a voice signal filtering models used to obtain the spectral shape of the signal frequency sound that is more subtle [5] [4].

2.4 Low-Pass Filter
Low pass filter is a type of filter that dampens high-frequency signals, usually used to carry the signal has a low frequency. The signal in question can be electrical signals or digital data such as images and sounds [9] [10].

2.5 High-Pass Filter
High pass filter is a type of filter that holds or lowers low frequencies, and capable of passing high frequencies. The workings of this high pass filter are by decreasing the output voltage for all frequencies that are below the cut-off frequency FC [9] [10].
2.6 Discrete Hartley Transform
Discrete Hartley Transform (DHT) and Discrete Fourier Transform (DFT) have the same characteristics [12]. Discrete Hartley Transform (DHT) [12], [13], [14], [15] plays an important role in many applications of digital signal processing (DSP) because it is a good alternative to the discrete Fourier transform (DFT) in the operation of real numbers. One of the main advantages of DHT is that it only involves different real computations with complex computations in DFT. The N-point DHT is given by:

\[ X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cos \left( \frac{2\pi nk}{N} \right) \]

where \( k = 0, 1 \ldots N - 1 \)

3. Research Method

3.1 Datasheet
The datasheet used in this research is the sound file with Chest Voice, Head Voice, Falsetto, and Vocal Fry voice registers obtained from the results of female voice recordings with the help of Adobe Audition 1.5 software.

3.2 Research Step
Step-designed research, in general, can be seen in Figure 1.

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Step-designed research in general [16],[17]

In figure 1 there are two processes, namely: the training process and the testing process. The training process the voice signal input is done pre-processing with pre-emphasis and continue to filter signals using low pass filter and a high pass filter results filter sound signal will be transformed using the Discrete Hartley Transform (DHT) and the transformation result is stored as a model pattern for the testing process, while in the testing process the voice signal input is processed by the pre-emphasize and continue to filter signals using a low pass filter and a high pass filter results filter sound signals
transformed using Discrete Hartley Transform (DHT), then enter the pattern matching mode that has been saved at the training stage, if the pattern model is similar or close to the training pattern, the output of the classification is a voice register. The overall research step in the introduction of the female voice register pattern constructed in this study can be illustrated in Figure 2.

4. Analysis and Discussion
4.1 Sample of Voice Register
The sample of voice registers used in research amounted to 138 samples which represent Chest Voice, Head Voice, Falsetto, and Vocal fry voice registers, 60% of 138 samples are used as training data and 40% are used as data testing, figure 3, 4, 5, 6 presents the frequency for each voice.

Figure 3. Voice register chest voice

Figure 4. Voice register head voice

Figure 5. Voice register falsetto

Figure 6. Voice register vocal fry
4.2 Pre-processing
In this research discuss two models of pre-processing by filtering the sound signal using a low pass filter and high pass filter, figure 8 illustrates the results of the low pass filter and figure 9 illustrates the results of the high pass filter.

![Figure 7. Examples of voice signals with Chest Voice voice registers](image)

![Figure 8. Result Low pass filter](image)

![Figure 9. Result High pass filter](image)

4.3 Testing
Testing voice register using Discrete Hartley Transform (DHT) conducted three experiments, results of the first experiment are presented in Table 1, the results of the second experiment are presented in Table 2 and the results of the third experiment are presented in Table 3.

| Table 1. First Test |
|---------------------|
|                     |
| **DHT**             |
| Chest Voice         | Head Voice | Falsetto | Vocalfry |
| Chest Voice         | 0.64       | 0.07     | 0.21     | 0.07     |
| Head Voice          | 0.08       | 0.58     | 0.08     | 0.25     |
| Falsetto            | 0.08       | 0.08     | 0.85     | 0        |
| Vocalfry            | 0.24       | 0        | 0        | 0.76     |
| **DHT + Low Pass Filter** |
| Chest Voice         | 0.64       | 0.07     | 0.21     | 0.07     |
| Head Voice          | 0.08       | 0.58     | 0.08     | 0.25     |
| Falsetto            | 0.08       | 0.08     | 0.85     | 0        |
| Vocalfry            | 0.24       | 0        | 0        | 0.76     |
| **DHT + High Pass Filter** |
| Chest Voice         | 0.64       | 0.07     | 0.21     | 0.07     |
| Head Voice          | 0.08       | 0.58     | 0.08     | 0.25     |
| Falsetto            | 0.08       | 0.08     | 0.85     | 0        |
| Vocalfry            | 0.24       | 0        | 0        | 0.76     |

| Table 2. Second Test |
|----------------------|
|                     |
| **DHT**             |
| Chest Voice         | Head Voice | Falsetto | Vocalfry |
| Chest Voice         | 0.57       | 0.14     | 0.14     | 0.14     |
| Head Voice          | 0.00       | 0.67     | 0.00     | 0.33     |
| Falsetto            | 0.00       | 0.00     | 0.85     | 0        |
| Vocalfry            | 0.12       | 0.18     | 0.06     | 0.65     |
| **DHT + Low Pass Filter** |
| Chest Voice         | 0.57       | 0.14     | 0.14     | 0.14     |
| Head Voice          | 0.00       | 0.67     | 0.00     | 0.33     |
| Falsetto            | 0.00       | 0.00     | 0.85     | 0        |
| Vocalfry            | 0.12       | 0.18     | 0.06     | 0.65     |
| **DHT + High Pass Filter** |
| Chest Voice         | 0.57       | 0.14     | 0.14     | 0.14     |
| Head Voice          | 0.00       | 0.67     | 0.00     | 0.33     |
| Falsetto            | 0.00       | 0.00     | 0.85     | 0        |
Vocalfry | 0.12 | 0.18 | 0.06 | 0.65

Table 3. Third Testing

|                      | Chest Voice | Head Voice | Falsetto | Vocalfry |
|----------------------|-------------|------------|----------|----------|
| DHT                  | Chest Voice | 0.64       | 0.21     | 0.14     | 0.00     |
|                      | Head Voice  | 0.08       | 0.58     | 0.17     | 017      |
|                      | Falsetto    | 0.15       | 0.15     | 0.69     | 0.00     |
|                      | Vocalfry    | 0.00       | 0.12     | 0.00     | 0.88     |
| DHT + Low Pass Filter| Chest Voice | 0.64       | 0.21     | 0.14     | 0.00     |
|                      | Head Voice  | 0.08       | 0.58     | 0.17     | 017      |
|                      | Falsetto    | 0.15       | 0.15     | 0.69     | 0.00     |
|                      | Vocalfry    | 0.00       | 0.12     | 0.00     | 0.88     |
| DHT + High Pass Filter| Chest Voice | 0.64       | 0.21     | 0.14     | 0.00     |
|                      | Head Voice  | 0.08       | 0.58     | 0.17     | 017      |
|                      | Falsetto    | 0.15       | 0.15     | 0.69     | 0.00     |
|                      | Vocalfry    | 0.00       | 0.12     | 0.00     | 0.88     |

In the first test well DHT, DHT with Pre-processing Low Pass Filter as well as DHT with Pre-processing High Pass Filter is able to identify 64% for the category of voice registers Chest Voice, 58% for Head Voice, 85% for falsetto and 76% for Vocal fry. The average results of the test on the first try at 71%.

In the second test well DHT, DHT with Pre-processing Low Pass Filter as well as DHT with Pre-processing High Pass Filter is able to identify 57% for the category of voice registers Chest Voice, 67% for Head Voice, 85% for falsetto and 65% for Vocal fry. The average results of the test on the second try at 68%.

In the third test well DHT, DHT with Pre-processing Low Pass Filter as well as DHT with Pre-processing High Pass Filter is able to identify 64% for the category of voice registers Chest Voice, 58% for Head Voice, 69% for falsetto and 88% for Vocal fry. The average results of the test on the third try at 70%.

In figure 10, figure 11, figure 12 clearly shows the results of the average testing of voice registers using DHT, DHT + High Pass Filter and DHT + Low Pass Filter did not obtain different classification results, from the three trials both DHT, DHT + High Pass Filter and DHT + Low Pass Filter is only able to classify voice registers with an average true positive rate of 69.67%.

5. Conclusions
The results showed that DHT was still not maximal in classifying voice register data, the use of Pre-processing filters both High Pass Filter and Low Pass Filter did not affect the results of the visit.
The sound frequency used is in ideal conditions so there is no noise that affects the classification results of voice registers.

6. References

[1] Fadlisyah and Muhathir, "Perbandingan Unjuk Kerja Transformasi Wavelet, Mellin Dan Discrete Sine Transform (Dst) Untuk Pengenalanayat Al-Qur'an Pada Surat Yasin 1-10 Melalui Suara," *Techsi-Jurnal Teknik Informatika*, vol. 7, no. 2, 2015.

[2] Sonali Goyal and Neera Batra, "Issues and Challenges of Voice Recognition in Pervasive Environment," *Indian Journal of Science and Technology*, vol. 10, no. 30, 2017.

[3] Suma Swamy and V Ramakrishnan K, "An Efficient Speech Recognition System," *Computer Science & Engineering: An International Journal (CSEIJ)*, vol. 3, no. 4, 2013.

[4] Himani Chauhan, Sumitra Samal, and Ankita Ghoshal, "VOICE RECOGNITION," *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol. 4, no. 4, 2015.

[5] Pal Singh Parwinder and Pushpa Rani, "An Approach to Extract Feature using MFCC," *IOSR Journal of Engineering (IOSRJEN)*, vol. 4, no. 8, 2014.

[6] R Meiyanti, A Subandi, N Fuqara, A Budiman M, and P U Siahaan A, "The Recognition Of Female Voice Based On Voice Registers In Singing Techniques In Real-Time Using Hankel Transform Method And Macdonald Function," in *International Conference On Computing And Applied Informatics*, 2017.

[7] Marek Fric, František Šram, and G Švec Jan, "Voice Registers, Vocal Folds Vibration Patterns And Their Presentation In Videokymography," in *International Acoustical Conference*, 2006.

[8] Morner M, Fransson F, and Fant G, "Voice register terminology and standard pitch," *Quarterly Progress and Status Report STL-QPSR*, 1963.

[9] P Hsu Hwei, *Schaum's Outlines Of Theory And Problems Of Signals And Systems.*: Mcgraw-Hill Companies, Inc, 1995.

[10] H Hayes Monson, *Schaum's Outlines Of Theory And Problems Of Signals And Systems.*: Mcgraw-Hill Companies, Inc, 1999.

[11] Kumar Singh Vinay, Shilpi Gupta, and D Dalal Upena, "Performance Comparison Of Discrete Hartley Transform (DHT) And Fast Fourier Transform (FFT) OFDM System In AWGN Channel," *International Journal Of Computer Applications*, vol. 70, no. 9, p. 2013.

[12] N Bracewell R, "Discrete Hartley transform," *Journal of the Optical Society of America*, vol. 73, no. 12, 1983.

[13] N Bracewell R, "The fast Hartley transform," in *Proceedings of the IEEE*, vol. 72, 1984.

[14] Narayan Murty M and K Sreekanth, "Relation Between Type-II Discrete Sine Transform And Type –I Discrete Hartley Transform," *Int. Journal Of Engineering Research And Application*, vol. 7, no. 6, 2017.

[15] Nikhil Singh and Anshuj Jain, "Review Paper On Discrete Hartley Transform Using Different Types Of Multiplier," *International Journal Of Innovative Research In Computer And Communication Engineering*, vol. 4, no. 10, 2016.

[16] Muhathir, "KLASIFIKASI EKSPRESI WAJAH MENGGUNAKAN BAG OF VISUAL WORDS," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING (JITE)*, vol. 1, no. 2, 2018.

[17] Muhathir, Herman Mawengkang, and Marwan Ramli, "KOMBINASI Z-FISHER TRANSFORM DAN BRAY CURTIS DISTANCE UNTUK PENGENALAN POLA HURUF JAR PADA CITRA AL-QURAN," *Jurnal Bisman Info*, vol. 4, no. 1, 2017.