

## PERBANDINGAN METODE PENGENALAN SINYAL BICARA BERBAHASA INDONESIA BERBASIS STATISTIK TERHADAP PENGENALAN SUARA BERBASIS FREKUENSI

Eva Damayanti<sup>1</sup>, Suhartono Tjondronegoro<sup>2</sup>, Koredianto Usman<sup>3</sup>

<sup>1</sup>Magister Elektro Komunikasi, Fakultas Teknik Elektro, Universitas Telkom

### Abstrak

Penelitian ini adalah perbandingan metode ekstraksi ciri berbasis statistik dan frekuensi untuk keperluan pengenalan suara. Penelitian ini bermaksud untuk menyelidiki efektivitas dari metode-metode tersebut.

Penelitian sistem identifikasi suara dilakukan dengan menguji sampel untuk mengidentifikasi suara orang yang berbeda. Langkah pertama yang dilakukan adalah pengambilan sampel suara yang memiliki format wav lalu disimpan ke dalam database. Suara yang telah disiapkan selanjutnya di ekstraksi cirinya dengan menggunakan metode ekstraksi ciri statistik atau frekuensi. Metode statistik yang dilakukan meliputi metode SSC, Autokorelasi dan LPC, sedangkan metode frekuensi terdiri dari FFT, STFT dan RSD. Hasil dari ekstraksi tersebut kemudian disimpan dalam vektor ciri dan dibandingkan dengan menggunakan euclidean distance. Nilai distance yang menunjukkan kesamaan ciri

Percobaan dilakukan pada suara laki-laki dan wanita yang mengucapkan huruf dan angka. Untuk huruf dibagi lagi atas kategori vokal dan konsonan.

Hasil yang diperoleh menunjukkan algoritma FFT untuk vokal wanita adalah yang paling baik untuk huruf dengan nilai distance tertinggi antar huruf adalah rata-rata 5,49 untuk konsonan dengan nilai distance tertinggi huruf adalah rata-rata 5,63 dengan algoritma FFT sedangkan untuk vokal laki-laki dengan nilai distance tertinggi adalah rata-rata 5,23 menunjukkan algoritma FFT begitu juga dengan huruf konsonan dengan nilai distance tertinggi adalah rata-rata 5,36, sedangkan untuk angka yang terbaik adalah FFT dengan distance rata-rata tertinggi 5,92

Kata Kunci : Pengenalan suara, statistik, frekuensi.

### Abstract

This study is aimed at investigating the effectiveness of two methods for speech recognition, the statistics-based feature extraction and the frequency-based feature extraction. The study of voice identification system was carried out by testing the samples of different human sounds

The process of obtaining and analyzing the voice has several steps. First, the voice samples in the wav format were collected and stored into the database. Second, the characteristics of those voices were extracted using statistical or frequency feature extraction methods. The statistical methods were conducted using the SSC method, autocorrelation and LPC, while the frequency method were carried out using FFT, STFT and RSD. Third, the results of the extraction were stored in a feature vector and then they were compared using euclidean distance, the distance value indicating the characteristics in common.

The experiments used male and female voices when they uttered letters, both vowels and consonants, and numbers. The results indicated that the best FFT algorithm was for the letters of female voices/vowels with the highest distance value among letters was 5.49 in average and the highest distance value for the consonant letters was 5.63 in average (with the FFT algorithm). The highest distance value of male voices/vowels was 5.23 in average (the FFT algorithm as well as) and the highest distance value for consonant was 5.36 in average. In addition, the best rate was the FFT with the highest average distance of 5.92.

Keywords : Voice recognition, statistics, frequency.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Suara merupakan suatu bentuk lazim dari sebuah komunikasi. Perkembangan terakhir menciptakan kemungkinan bahwa suara dapat digunakan dalam sistem keamanan<sup>[1]</sup>. Dalam pengenalan suara, tujuan yang ingin dicapai adalah menggunakan sampel dari suara untuk menentukan identitas seseorang yang mengeluarkan suara tersebut. Teknik ini memungkinkan penggunaan suara seseorang untuk memverifikasi identitas dan mengendalikan akses terhadap suatu layanan tertentu misalnya panggilan suara (*voice dialing*), *telepon banking*, *akses database*, *akses informasi*, bahkan akses ke suatu tempat/ruangan tertentu<sup>[6]</sup>.

Perkembangan riset pengenalan suara semakin meningkat, untuk hasil yang lebih baik setidaknya diketahui terlebih dahulu bahwa suara tersebut berasal dari seorang pria atau wanita. Pada penelitian sebelum sudah dilakukan penelitian tentang ekstraksi ciri untuk pengenalan sinyal suara yang digunakan untuk keperluan mengenali sinyal suara untuk menguji kemiripan sinyal suara atau dapat membedakan sinyal suara yang memiliki intonasi yang tinggi, rendah atau datar serta sinyal suara tanpa emosi maupun dengan emosi. Ekstraksi ciri yang digunakan pada penelitian sebelumnya adalah *Mel-Frequency Cepstrum Coefficients* (MFCC), *energy bit*, dan *Vector Quantization* (VQ)<sup>[11]</sup>. Metode ekstraksi ciri yang digunakan untuk penelitian ini adalah metode *berbasis Statistik dan berbasis frekuensi*. Di mana sinyal suara terlebih dulu akan melalui proses *preprocessing*, ekstraksi ciri berbasis statistik yaitu dengan *Statistical Signal Characterization* (SSC), autokorelasi dan *Linear Predictive Coding* (LPC), sedangkan pada ekstraksi ciri berbasis frekuensi yaitu dengan *Fast Fourier Transform* (FFT), *Short-Time Fourier Transform* (STFT) dan *Rapat Spektral Daya* (RSD). Dalam penerapannya, penelitian ini memanfaatkan MATLAB 2009a untuk mengolah data masukan yang didapatkan.

## 1.2 Tujuan

Tujuan dan manfaat yang ingin dicapai dalam penyusunan penelitian ini adalah :

1. Membandingkan algoritma pengenalan suara berbasis statistik dan frekuensi.
2. Mendapatkan kesimpulan pada setiap pengenalan suara baik yang berbasis statistik maupun berbasis frekuensi.
3. Mendapatkan parameter mana yang baik pada pengolahan berbasis statistik maupun berbasis frekuensi.

## 1.3 Perumusan Masalah

Beberapa permasalahan pada penelitian ini dapat diformulasikan sebagai berikut.

1. Bagaimana pengenalan suara berbasis statistik memakai parameter *Statistical Signal Characterization* (SSC), autokorelasi dan *Linear Predictive Coding* (LPC).
2. Bagaimana pengenalan suara berbasis frekuensi memakai parameter *Fast Fourier Transform* (FFT), *Short-Time Fourier Transform* (STFT) dan Rapat Spektral Daya (RSD).

## 1.4 Batasan Masalah

Untuk memaksimalkan kinerja penelitian yang dibuat serta menghindari terjadinya kesalahan persepsi mengenai fokus dan isi penelitian. Dalam penelitian ini dilakukan pembatasan masalah sebagai berikut :

1. Sumber sinyal suara latihan maupun uji adalah rekaman sinyal suara WAV, dengan yang menggunakan frekuensi sampling 8.000 Hz dan durasi  $\pm 2$  detik.
2. Jumlah rekaman sinyal suara untuk setiap orang terdiri dari angka (0-9) dan huruf (a-z) dalam bahasa Indonesia.
3. Pembuatan program dilakukan dengan menggunakan bantuan bahasa pemrograman Matlab seri 2009a.

## 1.5 Metodologi

Metodologi penelitian dalam penyusunan tesis ini meliputi :

---

1. Studi Literatur

Tahapan ini bertujuan untuk mengumpulkan bahan yang berupa buku referensi, artikel-artikel, jurnal-jurnal yang mendukung, dan dasar teori yang kuat tentang Ekstraksi Ciri Suara dan metode-metode apa saja yang dapat digunakan.

2. Analisa Desain

Tahap ini meliputi analisis kebutuhan untuk menganalisa rancangan untuk perbandingan dalam setiap algoritma di setiap metode

3. Implementasi Sistem

Tahap ini meliputi pembangunan perangkat lunak yang telah dirancang sebelumnya

4. Testing dan Analisis

Pada tahap ini dilakukan pengujian terhadap perangkat lunak yang telah dibangun sekaligus melakukan analisis terhadap hasil perangkat lunak yang berupa hasil analisis

5. Penyusunan Laporan

Pada tahap ini dilakukan penyusunan laporan hasil penelitian yang telah dilakukan dan membuat kesimpulan dari hal penelitian tersebut.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 KESIMPULAN

1. Hasil perhitungan jarak *Euclidean* pengenalan suara berbasis statistik menunjukkan bahwa parameter SSC (*Statistical Signal Characterization*) memberikan jarak terkecil, sedangkan parameter autokorelasi memberikan jarak terbesar. Artinya parameter autokorelasi lebih bisa membedakan suara, jadi dengan perkataan lain parameter autokorelasi lebih baik dari parameter SSC untuk dipakai pada sistem pengenalan suara.
2. Hasil perhitungan jarak *Euclidean* pengenalan suara berbasis frekuensi menunjukkan bahwa parameter RSD (*Rapat Spektral Daya*) memberikan jarak terkecil, sedangkan parameter FFT (*Fast Fourier Transform*) memberikan jarak terbesar. Artinya parameter FFT lebih bisa membedakan suara, jadi dengan perkataan lain parameter FFT lebih baik dari parameter RSD untuk dipakai pada sistem pengenalan suara.
3. Parameter yang baik adalah berbasis frekuensi, karena hasil jarak *Euclidean* pengenalan suara jaraknya terbesar (5,914) terhadap berbasis statistik (1,43).

#### 5.2 SARAN

1. Untuk penelitian selanjutnya, proses klasifikasi perlu dilakukan untuk memperkuat kesimpulan yang didapat pada penelitian ini.

## CONCLUSIONS AND RECOMMENDATIONS

---

### BAB V

## CONCLUSIONS AND RECOMMENDATIONS

### 5.1 CONCLUSION

Comparing the result of statistic and frequency, it can be concluded that as follows :

1. The result of Euclidean distance calculation using speech-base recognition statistics showed that the parameters of SSC ( Statistical Signal Characterization ) had the smallest distance whereas, the autocorrelation parameter had the greatest distance. This indicated that autocorrelation parameter was able to distinguish the sound better than SSC. In other words, for speech recognition system, using autocorrelation parameters result better than using SSC parameter.
2. The results of Euclidean distance calculation using speech-based recognition showed that the frequency of RSD parameters (Power Spectral Meeting) had the smallest distance; whereas the parameters FFT (Fast Fourier Transform) had the greatest distance. This indicated that FFT parameters were able to distinguish the sound better than RSD parameters. In other words, in speech recognition system the application of FFT parameters were better than the RSD parameters.
3. For the purpose of speech recognition, using frequency-based parameters is good, because this frequency-based parameters had the largest Euclidean distance of voice recognition (5.914) compared to the statistic-based (1.43) .

### 5.2 SUGGESTION

1. To validate the results of this study, another research applying classification process is necessary to be conducted.

**DAFTAR PUSTAKA**

- [1] *Analysis of a wav line*. <http://www.ymec.com/hp/signal2/gitar1.htm> di down load pada 26 Juli 2012 pukul 11.08 WIB
- [2] Furui, S., *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, Inc., New York, 1989.
- [3] Gerhard, D., *Pitch Extraction and Fundamental Frequency: History and Current Techniques*, Department of Computer Science University of Regina, Regina, 2003.
- [4] Hansel, D., B Littlefield, J Edyanto, *Matlab Bahasa Komputasi Teknis*, Andi, Yogyakarta, 2001.
- [5] Harinaldi, *Prinsip-Prinsip Statistik Untuk Teknik dan Sains*, Penerbit Erlangga, Jakarta, 2005.
- [6] Hidayatno, A. dan Sumardi, *Pengembangan dan Pendikte Otomatis Ucapan Bahasa Indonesia*, Fakultas Teknik UNDIP, Semarang, 2006.
- [7] Ifeachor, E.C., *Digital Signal Processing : A Practical Approach*, Addison-Wesley Publishers Ltd., New York, 1993.
- [8] Proakis, J., G., Manolakis D., G., *Pemrosesan Sinyal Digital: Prinsip, Algoritma, dan Aplikasi, Edisi bahasa Indonesia*, PT Prehallindo, Jakarta, 1997.
- [9] Rabiner, L., Biing-Hwang Juang. *Fundamentals Of Speech Recognition*, Prentice Hall, New Jersey, 1993.
- [10] Sudjana, *Metoda Statistika*, Penerbit Tarsito, Bandung, 2002.
- [11] Fachrudin AN (L2F099604), *Pengenalan Pengucap Tak Bergantung Teks dengan Metode Vector Quantization (VQ) Melalui Ekstraksi Linear Predictive Coding (LPC)*, Universitas Diponegoro

Telkom  
University