

PENCOCOKAN POLA SUARA (*SPEECH RECOGNITION*) DENGAN ALGORITMA FFT DAN *DIVIDE AND CONQUER*

Gressia Melissa – NIM 13506017

Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung
Jalan Ganesha nomor 10 Bandung
e-mail: if16017@students.if.itb.ac.id , ghe_gressia@yahoo.com

ABSTRAK

Voice recognition dibagi menjadi dua jenis, yaitu *speech recognition* dan *speaker recognition*. *Speech recognition* adalah proses identifikasi suara berdasarkan kata yang diucapkan. Parameter yang dibandingkan ialah tingkat penekanan suara yang kemudian akan dicocokkan dengan *template database* yang tersedia. Sedangkan sistem pengenalan suara berdasarkan orang yang berbicara dinamakan *speaker recognition*. Pada makalah ini hanya akan dibahas mengenai *speech recognition* karena kompleksitas algoritma yang diimplementasikan lebih sederhana daripada *speaker recognition*.

Algoritma yang akan diimplementasikan pada bahasan mengenai proses *speech recognition* ini adalah algoritma FFT (*Fast fourier transform*), yaitu algoritma yang cukup efisien dalam pemrosesan sinyal *digital* (dalam hal ini suara) dalam bentuk diskrit. Algoritma ini mengimplementasikan algoritma *Divide and Conquer* untuk pemrosesannya. Konsep utama algoritma ini adalah mengubah sinyal suara yang berbasis waktu menjadi berbasis frekuensi dengan membagi masalah menjadi beberapa upa masalah yang lebih kecil. Kemudian, setiap upa masalah diselesaikan dengan cara melakukan pencocokan pola *digital* suara.

Kata kunci: *divide and conquer*, FFT, identifikasi, perbandingan, *recognition*, *speech*, spektrum, suara, transformasi.

1. PENDAHULUAN

Biometrik, termasuk di dalamnya *speech recognition*, secara umum digunakan untuk identifikasi dan verifikasi. Identifikasi ialah mengenali identitas seseorang, dilakukan perbandingan kecocokan antara data *biometric*

seorang dalam *database* berisi record karakter seseorang. Sedangkan verifikasi adalah menentukan apakah seseorang sesuai dengan apa yang dikatakan terhadap dirinya.

Biometric recognition merupakan sistem pengenalan atau identifikasi seseorang berdasarkan karakteristik biologis khusus yang dimiliki oleh orang tersebut. Fungsinya selain untuk sistem keamanan dengan mengenali identitas seseorang, juga untuk identifikasi penyakit yang diderita seseorang, keperluan militer, dan lain-lain. Aplikasi *biometric recognition* antara lain retinal scan (identifikasi berdasarkan pola pembuluh darah pada retina mata), *fingerprint recognition* (identifikasi pola sidik jari unik pada setiap orang), *face recognition* (pengenalan seseorang berdasarkan raut dan ekspresi seseorang dengan kunci utama pada letak mata dan mulut), dan *voice recognition*.

Sedangkan *voice recognition* sendiri dibagi menjadi dua jenis, yaitu *speech recognition* dan *speaker recognition*. Berbeda dengan *speaker recognition* yang merupakan pengenalan identitas yang diklaim oleh seseorang dari suaranya (siri khusus dapat berupa intonasi suara, tingkat kedalaman suara, dan sebagainya), *speech recognition* adalah proses yang dilakukan komputer untuk mengenali kata yang diucapkan oleh seseorang tanpa mempedulikan identitas orang terkait. Implementasi *speech recognition* misalnya perintah suara untuk menjalankan aplikasi komputer.

Algoritma FFT (*Fast fourier transform*) merupakan salah satu metode untuk transformasi sinyal suara menjadi sinyal frekuensi. Artinya proses perekaman suara disimpan dalam bentuk *digital* berupa gelombang spektrum suara berbasis frekuensi.

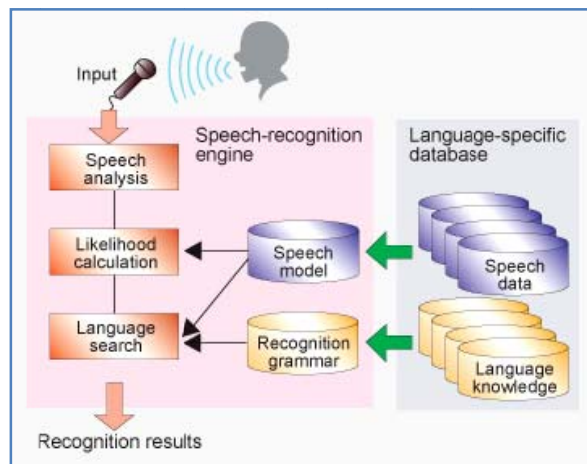
Sedangkan algoritma implementasi algoritma *divide and conquer* terletak pada pembagian objek permasalahan (data *digital*) menjadi upa masalah yang diselesaikan secara rekursif dan kemudian menggabungkan solusi masing-masing upa masalah sehingga membentuk solusi masalah semula pada tahap akhir.

2. IDENTIFIKASI POLA

2.1 Skema Utama *Speech Recognition*

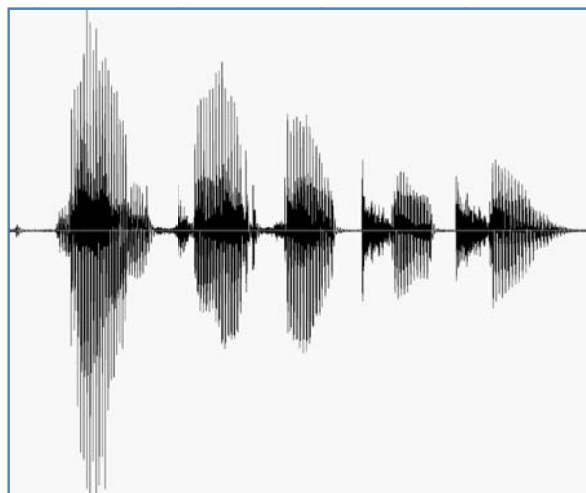
Terdapat 4 langkah utama dalam sistem pengenalan suara :

- [1] Penerimaan data input
- [2] Ekstraksi, yaitu penyimpanan data masukan sekaligus pembuatan *database* untuk *template*.
- [3] Perbandingan / pencocokan, yaitu tahap pencocokan data baru dengan data suara (pencocokan tata bahasa) pada *template*.
- [4] Validasi identitas pengguna.



Gambar 1. Skema *Speech Recognition*

2.2 Proses Pencocokan Pola Suara



Gambar 2. Spektrum Suara

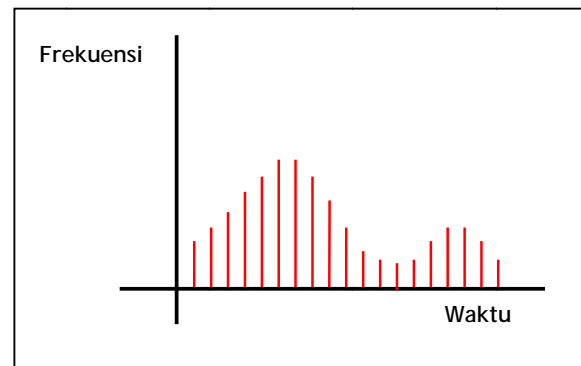
Secara umum, *speech recognizer* memproses sinyal suara yang masuk dan menyimpannya dalam bentuk *digital*. Hasil proses digitalisasi tersebut kemudian dikonversi dalam bentuk spektrum suara yang akan dianalisa dengan membandingkannya dengan *template* suara pada *database* sistem. Sebelumnya, data suara masukan dipilah-pilah dan diproses satu per satu berdasarkan urutannya. Pemilahan ini dilakukan agar proses analisis dapat dilakukan secara paralel.

Proses yang pertama kali dilakukan ialah memproses gelombang kontinu spektrum suara ke dalam bentuk diskrit. Langkah berikutnya ialah proses kalkulasi yang dibagi menjadi dua bagian :

- [1] Transformasi gelombang diskrit menjadi *array* data.
- [2] Untuk masing-masing elemen pada *array* data, hitung “ketinggian” gelombang (frekuensi).

Objek permasalahan yang akan dibagi adalah masukan berukuran n , berupa data diskrit gelombang suara.

Ketika mengkonversi gelombang suara ke dalam bentuk diskrit, gelombang diperlebar dengan cara memperinci berdasarkan waktu. Hal ini dilakukan agar proses algoritma selanjutnya (pencocokan) lebih mudah dilakukan. Namun, efek buruknya ialah *array of array* data yang terbentuk akan lebih banyak.



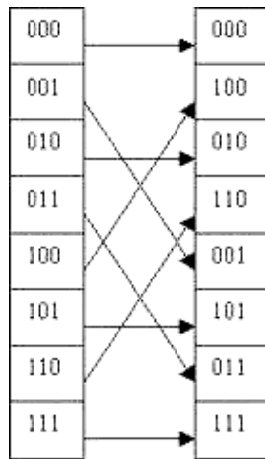
Gambar 3. Contoh Hasil Konversi Sinyal Diskrit

Dari tiap elemen *array* data tersebut, dikonversi ke dalam bentuk bilangan biner. Data biner tersebut yang nantinya akan dibandingkan dengan *template* data suara.

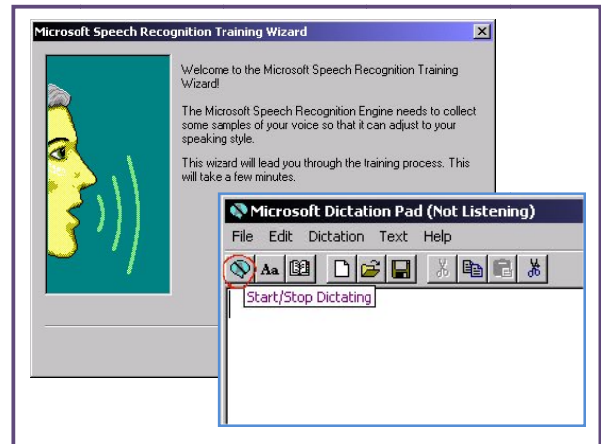
Proses *divide and conquer* :

- [1] Pilih sebuah angka N , dimana N merupakan bilangan bulat kelipatan 2. Bilangan ini berfungsi untuk menghitung jumlah elemen transformasi FFT.
- [2] Bagi dua data diskrit secara (dengan menerapkan algoritma *divide and conquer*) menjadi data diskrit yang lebih kecil berukuran $N = N_1 \cdot N_2$.
- [3] Upa masalah (objek data) dimasukkan ke dalam tabel (sebagai elemen tabel).

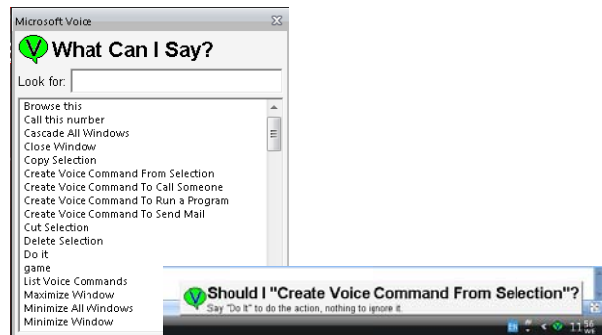
- [4] Untuk setiap elemen data, dicocokkan dengan data pada *template* (pada data *template* juga dilakukan pemrosesan digitalisasi menjadi data diskrit, dengan cara yang sama dengan proses digitalisasi data masukan baru yang ingin dicocokkan).
- [5] Setiap upa masalah disatukan kembali dan dianalisis secara keseluruhan, kecocokan dari segi tata bahasa dan apakah data yang diucapkan sesuai dengan kata yang tersedia pada *template* data.
- [6] Verifikasi data. Jika sesuai, proses lebih lanjut, sesuai dengan aplikasi yang mengimplementasikan algoritma ini.



Gambar 4. Perbandingan Array Data dengan Template



Gambar 5. Microsoft Dictation



Gambar 6. Microsoft Voice

3. APLIKASI

Terdapat beberapa pilihan yang dapat dilakukan selanjutnya, sesuai dengan aplikasi yang diimplementasikan, di antaranya ialah *voice command*. Komputer akan melakukan perintah sesuai dengan masukan pengguna. Contohnya pada aplikasi Microsoft *Voice*, ketika pengguna mengatakan “Start Notepad” dengan intonasi dan tata bahasa yang sesuai, komputer akan segera melakukan proses seperti yang telah dijelaskan pada bab sebelumnya. Jika sesuai dengan daftar perintah yang tersedia, aplikasi akan memastikan kata dengan menampilkan tulisan “Do you asked me to ‘Start Notepad’?”. Untuk memverifikasinya, pengguna cukup mengatakan “Do it”. Dan lagi-lagi proses penerimaan suara dilakukan seperti yang telah dibahas pada bab sebelumnya.

Aplikasi lainnya adalah Microsoft Dictation yang merupakan aplikasi *speech-to-text*, yaitu menuliskan apa yang diucapkan oleh pengguna. Jika dikembangkan lebih lanjut, software ini akan sangat berguna bagi orang-orang yang mempunyai keterbatasan fisik atau gerak.

4. KESIMPULAN

Speech recognition merupakan salah satu jenis *biometric recognition*, yaitu proses komputer mengenali apa yang diucapkan seseorang berdasarkan intonasi suara yang dikonversi ke dalam bentuk *digital print*.

Pengenalan pola suara adalah salah satu aplikasi yang berkembang saat ini. Sistem ini mengijinkan kita untuk berkomunikasi antara manusia dengan memasukkan data ke komputer. Salah satu fungsinya ialah untuk meningkatkan efisiensi industri manufaktur, mengontrol mesin dengan berbicara pada mesin itu.

Algoritma yang diimplementasikan untuk masalah pengenalan suara ini adalah algoritma *divide and conquer*. Proses awalnya ialah mengkonversi data spektrum suara ke dalam bentuk *digital* dan mengubah dalam bentuk diskrit.

Dari bentuk diskrit itulah yang kemudian mengaplikasikan algoritma *divide and conquer* untuk

mengoptimalkan waktu pencocokan (kompleksitas berkurang).

Contoh aplikasi yang menerapkan *speech recognition* antara lain *Microsoft Voice* dan *Microsoft Dictation*.

REFERENSI

- [1] Munir, Rinaldi, "Diktat Kuliah IF2251 Strategi Algoritmik", Program Studi Teknik Informatika ITB, 2007.
- [2] http://agusza.its-sby.edu/kuliah/citra/bab4_detail.html. Diakses pada tanggal 19 Mei 2008, pukul 15.05
- [3] <http://www.dspguru.com/info/faqs/ffifaq.htm>. Diakses pada tanggal 20 Mei 2008, pukul 14.00
- [4] <http://mathworld.wolfram.com/FastFourierTransform.html>. Diakses pada tanggal 19 Mei 2008, pukul 15.30
- [5] http://en.wikipedia.org/wiki/Spectrum_analyzer. Diakses pada tanggal 20 Mei 2008 pukul 12.15
- [6] http://en.wikipedia.org/wiki/Speaker_recognition. Diakses pada tanggal 15 Mei 2008 pukul 20.10
- [7] <http://www.relisoft.com/Science/Physics/fft.html>. Diakses pada tanggal 20 Mei 2008, pukul 14.30