

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan *speech synthesizer* beberapa dekade terakhir meningkat cukup pesat. Berawal dari hasil suara yang didapatkan tidak terdengar alami sama sekali, hingga menuju ke arah sintesis ucapan yang semakin baik. Salah satu contoh penerapan *speech synthesizer* adalah pada *Text-to-Speech*. *Speech synthesizer* merupakan sebuah sistem yang mampu menghasilkan suara tiruan manusia dengan sintesis ucapan. Terdapat tiga metode dasar yang dapat digunakan pada *speech synthesizer* yaitu *formant synthesis*, *articulatory synthesis* dan *concatenative synthesis*. Tugas Akhir ini akan membahas penyusunan unit ucapan menggunakan *diphone concatenation synthesis*. *Diphone* dipilih sebagai unit ucapan karena *diphone* merupakan pelebaran titik tengah dari bagian *phone* yang stabil ke titik tengah *phone* berikutnya. Oleh karena itu, tiap *diphone* mengandung transisi antar dua bunyi berdekatan (*adjacent phones*) yang stabil sehingga dapat mengurangi distorsi saat perangkaian antar *diphone*. Selanjutnya, *concatenation synthesis* merupakan metode perangkaian segmen-segmen ucapan yang telah direkam sebelumnya. Metode *concatenation* dianggap sebagai salah satu metode yang sesuai untuk membuat sintesis ucapan berkualitas tinggi karena hasil keluaran sistem berupa sinyal suara sintesis yang terdengar alami dan dapat dimengerti dengan jelas.

Langkah awal dalam membentuk sintesis ucapan adalah dengan merekam berbagai macam kata dalam Bahasa Indonesia dan disimpan pada *database*. Kemudian, hasil rekaman tersebut disegmentasi menjadi unit-unit *diphone*. Perangkaian unit-unit *diphone* dilakukan dengan menggunakan algoritma *Waveform Similarity Overlap-Add (WSOLA)*. Algoritma WSOLA memperkenankan menggunakan suara rekaman untuk dirangkai dengan lancar dan halus. Dengan menggunakan algoritma tersebut, perangkaian antar *diphone* yang mengandung transisi antar dua bunyi berdekatan diharapkan menjadi lebih halus dan dapat mengurangi bunyi yang bersifat letupan sehingga sintesis ucapan yang dihasilkan mampu memenuhi tiga kriteria, yaitu terdengar jelas dan dapat dimengerti, lancar dan alami (*intelligibility, fluidity, naturalness*).

1.2 Rumusan Masalah

Permasalahan yang diteliti dalam dalam tugas akhir ini adalah :

- 1) Bagaimana merancang suatu sistem *speech synthesizer* berbasis *diphone* dengan menggunakan metode *concatenation synthesizer*.
- 2) Bagaimana menerapkan algoritma *Waveform Similarity Overlap-Add* (WSOLA) sebagai teknik untuk menghaluskan perangkaian *diphone*.
- 3) Bagaimana menggunakan *database diphone* yang lengkap agar dapat melakukan perangkaian *diphone* dengan optimal.

1.3 Batasan Masalah

Beberapa permasalahan yang dibatasi adalah :

- 1) Metode *speech synthesizer* yang digunakan adalah *concatenation synthesizer* dengan unit ucapan *diphone*.
- 2) Algoritma yang diterapkan adalah *Waveform Similarity Overlap-Add* (WSOLA).
- 3) Format *diphone* sudah ditentukan dalam bentuk *.wav.
- 4) Kata yang diucapkan dalam Bahasa Indonesia dan mengacu pada Kamus Besar Bahasa Indonesia.
- 5) Persajakan (prosodi) tidak diperhatikan, sehingga suara yang keluar akan terdengar tanpa intonasi/datar.
- 6) Sistem hanya ditujukan untuk mengucapkan satu kata dalam Bahasa Indonesia.
- 7) *Database diphone* yang digunakan dalam aplikasi ini telah dibuat oleh Aggie Y Prihandi.

1.4 Tujuan dan Manfaat Penelitian

Adapun tujuan penyusunan tugas akhir ini adalah :

- 1) Meneliti dan merancang sistem *speech synthesizer* berbasis *diphone* dengan menggunakan metode *concatenation*.
- 2) Mempelajari dan menganalisis performansi yang dihasilkan oleh penerapan algoritma WSOLA yang digunakan untuk perangkaian *diphone*.

- 3) Menggunakan *database diphone* yang lengkap sebagai pendukung sistem agar didapatkan sintesis suara yang optimal.

Manfaat yang diharapkan pada penyusunan tugas akhir ini antara lain :

- 1) Dapat membuat suatu sistem *speech synthesizer* sederhana dengan menggunakan algoritma WSOLA.
- 2) Dapat menjadikan transisi antara dua *diphone* yang berdekatan menjadi lancar dan meminimalkan bunyi yang sifatnya eksplosif pada daerah perangkaian.
- 3) Mendapatkan hasil sintesis ucapan yang optimal, alami dan dapat dimengerti dari metode ini melalui simulasi dan pengujian yang dilakukan.

1.5 Metodologi Penelitian

Langkah – langkah yang digunakan dalam pengerjaan Tugas Akhir ini adalah :

- 1) Studi literatur

Langkah ini dilaksanakan dalam bentuk :

- a. Mempelajari karakter-karakter *speech* seperti *pitch*, frekuensi *formant* dan energi
- b. Mempelajari metode *diphone concatenation synthesizer*
- c. Mempelajari konsep algoritma WSOLA

- 2) Perancangan Sistem

Perancangan sistem berdasarkan algoritma yang telah dipelajari dan menyesuaikan dengan bahasa pemrograman yang digunakan.

- 3) Pengujian dan analisis

Langkah ini terdiri dari :

- a. Menguji kemampuan algoritma WSOLA untuk menggabungkan unit-unit *diphone* dari hasil rekaman dengan menggunakan software MATLAB R2008a
- b. Menganalisis dan menyimpulkan hasil sintesis ucapan yang terdengar berdasarkan perangkaian unit-unit *diphone* yang dilakukan dengan algoritma WSOLA
- c. Penyusunan laporan tugas akhir dan kesimpulan akhir

1.6 Hipotesis

Pengerjaan tugas akhir ini diawali dengan menyusun hipotesis sebagai berikut :

- 1) Sintesis ucapan berbasis *diphone* dengan algoritma WSOLA akan memberikan hasil yang lancar, terdengar alami dan dapat dimengerti.
- 2) Semakin beranekaragam unit-unit *diphone* yang terdapat pada *database* akan memudahkan dalam pembentukan sintesis ucapan, namun semakin besar memori yang dibutuhkan untuk menyimpan unit-unit *diphone* tersebut.

1.7 Sistematika Penulisan

Tugas akhir ini disusun dalam lima bab, yaitu :

- I. BAB I : Pendahuluan**
Berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metodologi penulisan dan sistematika penulisan.
- II. BAB II : Dasar Teori**
Berisi tentang teori yang mendukung dan mendasari penulisan tugas akhir ini, yaitu tentang teori dasar *speech*, *pitch*, *formant*, *speech synthesis*, *concatenation synthesis* dan algoritma *Waveform Similarity Overlap-Add (WSOLA)*.
- III. BAB III : Perancangan Sistem**
Berisi perancangan *concatenation synthesizer* dimana unit-unit *diphone* hasil rekaman suara yang ada pada *database* digabungkan dengan menggunakan algoritma WSOLA.
- IV. BAB IV : Analisis Hasil Pengujian**
Berisi analisis dari hasil simulasi mengenai sintesis ucapan berbasis *diphone* dengan menggunakan algoritma WSOLA.
- V. BAB V : Kesimpulan dan Saran**
Berisi kesimpulan dari analisis yang dilakukan dan saran untuk pengembangan lebih lanjut.