

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi bahasa adalah teknologi yang berhubungan dengan penggunaan bahasa, baik bahasa lisan maupun bahasa tulisan. Bahasa merupakan alat komunikasi paling baik dan tepat sasaran untuk menyampaikan keinginan dan maksud manusia. Bentuk representasinya adalah berupa suara atau ucapan (*spoken language*), tetapi sering pula dinyatakan dalam bentuk tulisan.

Sistem konversi *text-to-speech* (TTS) merupakan suatu sistem yang mampu memproduksi sinyal ucapan secara otomatis melalui transkripsi *grafem-ke-fonem* untuk kalimat yang diucapkan [7]. Dalam bidang kajian pengolahan sinyal ucapan (*speech processing*), salah satu yang banyak dikaji dan diteliti adalah mengenai pembuatan pensintesis sinyal ucapan (*speech synthesis*), khususnya mengenai pengkonversian teks ke ucapan *text to speech* (TTS). Sistem ini nantinya sangat berguna sebagai *human machine communication*.

Penelitian TTS yang sudah pernah dilakukan adalah dengan menggunakan *neural network* jenis *multirate recurrent neural network*. Dari hasil penelitian terdapat empat aspek. Dua aspek menunjukkan hasil yang sudah baik, antara lain sistem TTS sudah mampu mengucapkan vokal /a/, /i/, /u/, /e/, /o/, dan /ə/ dengan benar, baik vokal pada posisi awal, tengah, akhir kata, maupun vokal serupa yang beriringan, selain itu juga terlihat adanya kemiripan pada bentuk gelombang sinyal ucapan sintesis dan sinyal ucapan asli jika dilihat dalam representasi domain waktu dan spektogramnya. Akan tetapi dua aspek lainnya masih menunjukkan hasil yang kurang, dimana diketahui bahwa untuk pengucapan pada rangkaian kalimat masih mendapatkan nilai RMSE yang besar dan koefisien korelasi yang kecil, artinya secara kualitatif intonasi dan durasi sinyal ucapan hasil sintesis masih memiliki kemiripan yang rendah dibandingkan dengan sinyal ucapan asli, selain itu diketahui juga bahwa pada beberapa

sambungan antar *fonem*, pensintesis masih belum bekerja dengan baik, dimana ditunjukkan oleh tidak halus nya titik persambungan yang kadang terdengar dalam ucapan sintesis sebagai klik/pop.

Yang menjadi fokus penelitian saat ini adalah ingin mendapatkan kualitas suara yang alami dengan pengucapan yang lebih baik dan lancar. Dalam hal ini bahwa proses pengubahan teks menjadi adalah sesuai dengan kaidah bahasa dan kaidah persajakan, sehingga hasil ucapan sintesis terdengar jelas dan dapat dimengerti, lebih lancar dan lebih alami (*intelligibility, fluidity, naturalness*).

1.2 Tujuan

Tujuan penulisan tugas akhir ini adalah:

1. Mengimplementasikan sistem TTS bahasa Indonesia yang mampu menghasilkan sinyal ucapan yang jelas dan alami dengan pembangkitan model prosodi menggunakan *Hopfield Neural Network*.
2. Merealisasikan model kontrol prosodi berbasis *Hopfield Neural Network*.

1.3 Perumusan Masalah

Permasalahan yang akan dibahas dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana merealisasikan ritme, tempo, aksen, intonasi dan penekanan yang tepat dengan kontrol prosodi berbasis *Hopfield Neural Network*.
2. Bagaimana membangkitkan sinyal ucapan berdasarkan informasi prosodi yang dihasilkan dari output *Hopfield Neural Network*.

1.4 Batasan Masalah

Beberapa hal yang menjadi batasan pada tugas akhir ini adalah sebagai berikut:

1. Hanya satu pembicara yang akan direkam ucapannya dan pemodelan intonasi sesuai dengan intonasi pembicara tersebut.
2. Bahasa target adalah Bahasa Indonesia yang sesuai dengan Ejaan Yang Disempurnakan.

3. Teks masukan terbatas pada bahasa baku dan karakter huruf serta tanda baca dan spasi, tidak mengakomodasi singkatan, bahasa serapan dan penulisan angka serta karakter lainnya.
4. Blok TTS tidak direalisasikan seluruhnya hanya terfokus pada bagian kontrol prosodi dan modul *synthesizer*.
5. Algoritma sintesis menggunakan *overlap-add* dengan *pitch* sinkron yang dimodifikasi.
6. Modifikasi *pitch* dan modifikasi waktu menggunakan metode sederhana.
7. Pemanfaatan kamus fonetik dan *database* rekaman sinyal ucapan hanya seperlunya saja dengan tetap mendukung penyelesaian masalah.
8. Simulasi dengan metode *Hopfield Neural Network* dibuat dengan menggunakan Matlab.
9. Arti teks masukan tidak dibahas dan tidak menjadi permasalahan.
10. Jumlah kalimat teks masukan dan jumlah kata dalam tiap kalimat akan dibatasi seiring kebutuhan sistem untuk tetap mendukung tercapainya penyelesaian masalah.
11. Modul *rule based algorithm* tidak direalisasikan.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam menyelesaikan tugas akhir ini adalah:

1. Studi literatur, mempelajari teori mengenai konversi teks ke ucapan, dan beberapa hasil penelitian yang telah dilakukan dalam beberapa jurnal yang berkaitan dengan pensintesis sinyal ucapan menggunakan *Hopfield Neural Network*.
2. Pendesainan sistem pensintesis sinyal ucapan.
3. Analisis unjuk kerja dari sistem yang telah dibuat.

1.6 Sistematika Penulisan

Tugas akhir ini dibagi dalam beberapa topik bahasan yang disusun secara sistematis sebagai berikut :

Bab I : Pendahuluan

Berisikan pembahasan mengenai latar belakang, tujuan penulisan, perumusan masalah, batasan masalah, metodologi penelitian serta sistematika penulisan.

Bab II : Dasar Teori

Berisikan penjelasan mengenai konsep dan teori secara singkat dari sistem pengkonversi teks ke ucapan (*teks-to-speech*), dan jaringan syaraf tiruan yang digunakan yaitu *Hopfield Neural Network* (HNN).

Bab III: Perancangan dan Realisasi Sistem

Berisikan proses perancangan dan realisasi pensintesis sinyal ucapan beserta penjelasan mengenai beberapa blok yang digunakan.

Bab IV: Analisis Hasil Pengujian

Berisikan hasil pengukuran, hasil pengujian dan analisis terhadap unjuk kerja system yang direalisasikan.

Bab V : Penutup

Memberikan kesimpulan hasil penelitian dan saran pengembangan penelitian berikutnya.