BAB I
PENDAHULUAN

1.1 Latar Belakang

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. Perbedaan sistem TTS dengan *talking machine* biasa adalah keotomatisannya dalam mengucapkan kata-kata baru. Oleh karena itu TTS memungkinkan untuk diimplementasikan pada bidang aplikasi yang beragam seperti aplikasi chatting bicara, sms bicara, buku digital dan pembaca email otomatis. Luasnya aplikasi yang ditawarkan oleh sistem TTS ini, dan berkembangnya beberapa perangkat/*platform*, seperti *Handphone* dan *PDA*, telah mendorong diimplementasikannya sistem TTS pada berbagai *platform* untuk berbagai keperluan. Dukungan *hardware* dan *software* yang memadai memungkinkan sistem TTS untuk diimplementasikan pada perangkat tersebut.

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, sintesiser masih belum bekerja dengan baik, dimana ditunjukkan oleh tidak halusnya titik persambungan yang kadang terdengar dalam ucapan sintesis sebagai klik/pop.

Agar tercapai sintesis ucapan yang alami, sangat membutuhkan kontrol prosodi untuk membangkitkan informasi linguistik yang terdapat dalam teks masukan, seperti memastikan pengaturan ritme, tempo, aksen, intonasi, dan penekanan yang tepat. Oleh karena melihat kondisi diatas maka permasalahan pengontrolan prosodi menjadi fokus dalam thesis ini. Sehingga diputuskan bagian-bagian yang akan dikerjakan pada thesis kali ini yaitu hanya terfokus pada modul kontrol prosodi dan modul synthesizer dengan tujuan tentu saja untuk memperbaiki hasil penelitian TTS sebelumnya. Sementara modul text analyzer tidak direalisasikan karena dari hasil penelitian sebelumnya yang telah dipaparkan di atas sudah menunjukkan hasil yang baik. Selain itu inovasi yang dilakukan adalah dengan menggunakan metode neural network yang lain yaitu Hopfield Neural Network, karena neural network jenis ini belum pernah dipakai untuk kasus TTS.

1.2 Perumusan Masalah

Permasalahan yang akan dibahas dalam thesis ini adalah sebagai berikut :

1. Bagaimana merealisasikan model kontrol prosodi berbasis HNN, hal ini terkait dengan modul ekstraksi ciri yang dikerjakan.
2. Bagaimana membangkitkan sinyal ucapan sintesis berdasarkan informasi prosodi yang diperoleh dari output network HNN, hal ini terkait dengan modul synthesizer atau modul generator intonasi prosodi yang dikerjakan.

1.3 Tujuan & Kegunaan

1.3.1 Tujuan

Tujuan penyusunan thesis ini adalah :

1. Untuk merealisasikan model kontrol prosodi berbasis HNN, hal ini terkait dengan modul ekstraksi ciri yang dikerjakan.
2. Untuk membangkitkan sinyal ucapan sintesis berdasarkan informasi prosodi yang diperoleh dari output network HNN, hal ini terkait dengan modul synthesizer atau modul generator intonasi yang dikerjakan.

1.3.2 Kegunaan

Kegunaan penyusunan thesis ini, sistem memungkinkan untuk diimplementasikan pada bidang aplikasi yang beragam yang berbasis concatenation synthesizer. Beberapa aplikasi dimana sistem ini dapat diterapkan antara lain :

- Layanan telekomunikasi seperti akses informasi tekstual melalui telepon, *Telephone relay service,* SMS bicara, chatting bicara
- Pembacaan e-mail melalui telepon, IVR (*Interactive Voice Respons*), dan *Automatic Caller Identity*
- Pembelajaran bahasa
- Pemecahan terhadap beberapa permasalahan kemanusiaan seperti membantu tuna netra mengakses informasi tertulis atau membantu tuna wicara mengungkapkan pesannya dalam bentuk ucapan
- Buku bicara (*talking books*) dan mainan bicara
- Multimedia, komunikasi *man-machine*
- Penelitian dasar dan terapan

1.4 Hipotesis

Pada penelitian sebelumnya telah berhasil direalisasikan sistem text to speech bahasa indonesia dengan menggunakan database fonem. Penelitian tersebut berbasis modifikasi pada domain frekuensi dimana dilakukan prediksi kontur F0 (frekuensi fundamental). Sedangkan pada penelitian thesis ini berbasis modifikasi pada domain waktu dimana dilakukan kontrol prosodi untuk menentukan ciri delta dari setiap suku kata dan tentu saja menggunakan database suku kata untuk mempermudah. Dalam thesis ini terdapat beberapa hipotesis yang diuji yaitu :

1. Semakin banyak vector ciri input yang ditraining maka waktu komputasi algoritma HNN akan semakin lama. Dan dengan kemampuan HNN dalam mengenali pola input yang terdistorsi

maka diprediksi hasil akurasi mampu mencapai 100%, oleh karena itu dalam thesis ini dipilih menggunakan metode HNN.

2. Metode ekstraksi ciri scenario dua yang memiliki panjang vector ciri lebih pendek diprediksi memerlukan waktu komputasi algoritma yang lebih cepat dan dapat lebih menghemat memori.
3. Modul generator intonasi yang diuji masih memungkinkan belum bisa mendapatkan nilai RMSE yang kecil dan nilai koefisien korelasi yang besar.
4. Dengan ciri delta yang disimpan maka sangat memungkinkan kalimat sintesis lebih berintonasi, maka dari itu dilakukan modifikasi pada domain waktu karena ciri delta yang disimpan merupakan informasi pada domain waktu..
5. Jika dilihat dari representasi domain waktu dan spektogram, maka masih bisa memungkinkan terlihat kemiripan antara bentuk gelombang sinyal ucapan sintesis dan sinyal ucapan asli.

1.5 Pembatasan Masalah

Beberapa hal yang dibatasi pada thesis ini adalah sebagai berikut :

1. Bahasa target adalah Bahasa Indonesia yang sesuai dengan ejaan yang disempurnakan.
2. Teks masukan hanya terbatas pada bahasa baku dan karakter huruf serta tanda baca dan spasi, tidak mengakomodasi singkatan, bahasa serapan dan penulisan angka serta karakter lainnya.

3. Blok TTS tidak direalisasikan seluruhnya, hanya fokus pada bagian modul kontrol prosodi dan modul synthesizer.
4. Pemanfaatan kamus fonetik dan database rekaman sinyal ucapan hanya seperlunya saja dengan tetap mendukung tercapainya penyelesaian masalah.
5. Simulasi dengan metode HNN dibuat dengan menggunakan Matlab.
6. Arti kalimat teks masukan tidak dibahas dan tidak menjadi permasalahan.
7. Jumlah kalimat teks masukan adalah satu kalimat yang terdiri dari tiga kata.
8. Modul algorithm rule based tidak direalisasikan.
9. Tidak dilakukan modifikasi pada domain frekuensi.

1.6 Metode Penelitian

Metodologi yang digunakan dalam menyelesaian thesis ini adalah :

1. Studi literatur, yaitu dengan mempelajari konsep dan teori-teori mengenai prosodic kontrol, sintesa ucapan, dan HNN.
2. Analisis kebutuhan sistem berdasarkan permasalahan yang ada.
3. Realisasi model kontrol prosodi sesuai yang diinginkan.
4. Perancangan simulasi kontrol prosodi dan synthesizer pada Matlab.

5. Pengujian model sistem yang didesain dengan masukan karakter kata atau kalimat serta memanipulasi variabel-variabel yang ada dan meneliti pengaruhnya terhadap kinerja sistem.
6. Pengambilan kesimpulan dan penyusunan laporan thesis.

BAB II
LANDASAN TEORI

2.1 Komponen Dasar Sistem Text-to-Speech

Sistem TTS memproduksi sinyal ucapan secara otomatis melalui transkripsi *grafem* ke *fonem* pada kalimat yang diberikan. Hal inilah yang membedakan sistem TTS dengan mesin bicara lainnya. Sistem *voice response sistems* misalnya, bekerja dengan merangkai susunan kata terpisah (*isolated word*), hanya sesuai untuk aplikasi dengan jumlah kosa kata yang terbatas. Dalam konteks sistem TTS, sangat tidak mungkin (selain tidak berguna) untuk menyimpan seluruh kata dari satu bahasa.

Tugas sistem TTS secara umum dapat dibagi dalam 2 bagian besar, analisa teks dan sintesa ucapan. Analisa teks mentransformasi teks masukan menjadi representasi linguistik, selanjutnya bagian sintesis ucapan mentransformasi representasi linguistik tersebut menjadi gelombang sinyal ucapan.

Secara umum proses dalam sistem TTS terdiri dari *Natural Language Prossesing* (NLP) yang berupa modul konversi teks ke *fonem* yang menghasilkan transkripsi fonetik beserta informasi intonasi dan ritme (dikenal dengan prosodi) dan *Digital Signal Processing* (DSP) yang berupa modul konversi *fonem* ke ucapan, yang mengubah