SCHEME OF TEXT TO SPEECH WITH MULTILAYER PERCEPTRON NEURAL NETWORK PROSODY MODEL AT MBROLA

Seandy Arandiant Rozano[1], M. Ramdhani[2], Iwan Iwut Tirtoasmoro[3]

[1]Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

[1]**09850067@imtelkom.ac.id**

Abstrak

Text to speech (TTS) merupakan salah satu aplikasi dalam bidang teknologi informasi sebagai salah satu cara interaksi antara manusia dan komputer dengan cara mengkoversi teks menjadi ucapan/suara. Saat ini sudah dibuat TTS Bahasa Indonesia untuk penggunaan di PC yaitu IndoTTS, namun pelafalan pada IndoTTS ini masih belum natural.

Dalam sistem Text to Speech (TTS), sebenarnya pembentukan intonasi yang benar merupakan faktor penting yang akan mempengaruhi pembacaan pada output sistem. Bagian yang mengatur pembentukan intonasi di bagian output sistem ini disebut pembangkitan prosodi. Intonasi yang dihasilkan dari bagian ini meliputi durasi dan pitch pengucapan teks input.

Proses penentuan nilai durasi dan pitch dari teks input bersifat kompleks dan non-linear, maka sebagai dasar sistem digunakan Multilayer Perceptron Neural Network (MLPNN) sebagai model prosodinya. Model prosodi berbasis MLPNN membentuk intonasi pengucapan teks input dengan cara menentukan nilai durasi dan pitch dari tiap fonem penyusun teks input. Penentuan nilai durasi dan pitch dilakukan setelah sistem melakukan pembelajaran terhadap sampel pengucapan dari suatu kalimat.

Dalam tugas akhir ini, telah dibuat aplikasi Text to Speech Bahasa Indonesia dengan disertai peningkatan natural (kealamian) dalam pelafalan kalimat berupa teks (tidak termasuk angka dan simbol-simbol). Pembuatan aplikasi TTS ini menggunakan bahasa pemrograman Borland Delphi 7.0 dan memanfaatkan database diphone Bahasa Indonesia yang sudah tersedia serta menggunakan pembangkit ucapan Mbrola. Setelah dilakukan pengambilan MOS dari 30 koresponden yang terdiri dari mahasiswa dan masyarakat sekitar, didapatkan bahwa hasil dari sistem TTS dengan model prosodi MLPNN lebih baik kualitasnya dibandingkan dengan sistem IndoTTS.

Kata Kunci : Prosodi, MLPNN, diphone, MBROLA.

Abstract

Text to Speech is one of Information Technology application which is used as an interaction between human and computer by converting text becoming voice. Nowadays, TTS Bahasa Indonesia has been made for PC use that is IndoTTS, but IndoTTS reading is still unnatural.

In Text to Speech system, a forming of the right intonation is very important factor that will influence voice in system output. Part of this system that arranges intonation forming in output part called 'prosody evocation'. Intonation, resulted from this part, includes duration and pitch of input text pronunciation.

Process of duration value determination and pitch from input text have complex and non-linear characteristic, so as a base of prosody model system Multilayer Perceptron Neural Network (MLPNN) is used. Prosody model based on MLPNN produces intonation of input text pronunciations by deciding duration value and pitch from every phoneme of input text compiler. Determination of duration value and pitch is done after system does some learning about pronunciation sample from a sentence.

In this final task, Text to Speech of Indonesian language application has be made and accompanied by natural improvement in reading a sentence without numeral and symbols. This application is done by using programming language Borland Delphi 7.0 with diphone Indonesian language database which has been available and MBROLA pronounciation generator. After getting the MOS result from 30 coresponden, it can be concluded that TTS system with MLPNN prosody model has better quality compared with IndoTTS system.

Keywords : Prosody, MLPNN, diphone, MBROLA.

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Text to Speech merupakan salah satu program aplikasi yang berbasis interaksi manusia dan komputer, yaitu aplikasi yang bekerja dengan cara mengkonversi teks *input* yang dimasukkan ke dalam program menjadi *output* berupa suara. Aplikasi tersebut sudah dikembangkan di Indonesia, yaitu IndoTTS, yang melakukan proses pengubahan teks *input* Bahasa Indonesia ke dalam bentuk ucapan.

Dalam *Text to Speech* pembentukan intonasi yang benar merupakan faktor penting yang akan mempengaruhi pembacaan pada *output* sistem yaitu dengan penambahan prosodi atau atribut linguistik terhadap fonem-fonem penyusun teks *input*, karena atribut linguistik merupakan faktor penentu keras lemahnya (intonasi) serta lamanya (durasi) pengucapan suatu fonem pada kata atau kalimat. Nilai durasi dan *pitch* memiliki relasi *non-linier* dengan fonem, tergantung kata atau konteks dari kalimat. Oleh karena itu diperlukan sistem prosodi agar dapat diperoleh nilai durasi dan *pitch* yang proporsional untuk menghasilkan ucapan yang natural, ucapan dengan nada yang *acceptable* untuk didengar.

1.2 Perumusan Masalah

Permasalahan yang akan dibahas dalam tugas akhir ini adalah pengembangan sistem prosodi untuk meningkatkan kualitas pengucapan kata atau kalimat yang dihasilkan dengan menggunakan model prosodi berbasis *Multilayer Perceptron Neural Network.*

1.3 Tujuan Penelitian

Tujuan penelitian tugas akhir ini adalah mengembangkan sistem *Text to Speech* yang ada dengan meningkatkan kualitas dari pengucapan pada *output*nya.

1.4 Batasan Masalah

Ruang lingkup penelitian yang dilakukan, yaitu :

1. Penelitian ini merupakan pengembangan dari sistem IndoTTS yang sudah ada.
2. Sistem yang dikembangkan adalah sistem prosodi dengan basis *MLPNN*, untuk meningkatkan kualitas dari pengucapan.
3. Bahasa pemrograman yang digunakan adalah Delphi 7.0
4. Menggunakan *Mbrola* sebagai converter fonem menjadi ucapan.
5. Aplikasi *Text to Speech* disesuaikan dengan spesifikasi IndoTTS dengan penggunaan *Mbrola* dan *database diphone* Id1.
6. Pengimplementasian sistem lebih difokuskan pada bagian *Natual Language Processing (NLP).*
7. Pengimplementasian sistem dibatasi dengan kalimat berupa teks saja (tidak termasuk angka dan simbol-simbol).
8. Analisa sistem dibatasi dengan 2 kata, 3 kata, dan 4 kata sesuai dengan sampel pada sistem.
9. Kejelasan pengukuran sistem dalam penentuan durasi dan *pitch* yang tepat, bersifat subjektif berdasarkan pada pengambilan *MOS* dari 30 koresponden.

1.5 Metodologi Penelitian

Metode penelitian yang digunakan adalah sebagai berikut :

1. Studi literatur.

Mengambil referensi-referensi yang menunjang penelitian tugas akhir ini seperti pemrograman Delphi 7.0, dasar-dasar teori *Text to Speech*, modeling prosodi pada *Text to Speech*, dan hal-hal yang berhubungan dengan tema tugas akhir ini baik dari buku, paper, maupun artikel ilmiah.

2. Pengembangan sistem

Mengembangkan sistem prosodi untuk mendapatkan kualitas pengucapan yang lebih baik pada otuputnya.

3. Pengujian dan analisa sistem

Pengujian sistem, dilakukan dengan memasukkan *input* ke sistem untuk diteliti kualitas ucapan yang dihasilkan. Lalu membandingkan hasil pengucapan sebelum dan sesudah pengembangan sistem.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang masalah yang ada, perumusan masalah, tujuan dari penelitian, batasan-batasan masalah, metodologi yang dipakai dalam penelitian, sistematika penulisan dan rencana kerja penyusunan tugas akhir ini.

BAB II LANDASAN TEORI

Bab ini menjelaskan dan merangkum teori-teori dasar yang menyangkut permasalahan yang dibahas.

BAB III PERANCANGAN dan IMPLEMENTASI SISTEM

Bab ini membahas proses pengembangan dari sistem prosodi.

BAB IV PENGUJIAN DAN ANALISA SISTEM

Bab ini membahas proses pengujian sistem serta analisis kinerja sistem setelah pengembangan.

BAB V KESIMPULAN DAN SARAN

Menyimpulkan seluruh hasil penelitian dan saran-saran untuk penelitian dan pengembangan berikutnya.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. Proses pembelajaran sistem *TTS* model prosodi *MLPNN* yang dikhususkan pada bagian NLP (*Natural Language Processing*) mampu menghasilkan ucapan yang memiliki intonasi yang cukup natural.
2. Hasil MOS pengujian kejelasan kata yang dihasilkan sistem *TTS* model prosodi *MLPNN* menunjukkan kualitas yang lebih baik daripada sistem IndoTTS. Sistem TTS model prosodi MLPNN mendapat nilai rata-rata sebesar 3.51 sedangkan sistem IndoTTS mendapat nilai rata-rata sebesar 3.023.
3. Hasil MOS pengujian kenaturalan kata yang dihasilkan sistem *TTS* model prosodi *MLPNN* menunjukkan kualitas yang lebih baik daripada sistem IndoTTS. Sistem TTS model prosodi MLPNN mendapat nilai rata-rata sebesar 3.24 sedangkan sistem IndoTTS mendapat nilai rata-rata sebesar 3.165.
4. Sistem IndoTTS masih kesulitan dalam melakukan pembacaan kata yang memiliki variasi fonem yang kompleks sehingga pembacaan mengalami *error*. Hal ini dapat diatasi oleh sistem *TTS* dengan prosodi *MLPNN*, karena pada sistem ini terlebih dahulu melakukan inisialisasi fonem-fonem saat pembelajaran, sehingga kata dengan fonem yang kompleks pun dapat dilafalkan.
5. Semakin banyak jumlah *training* yang dilakukan pada sistem *TTS* dengan model prosodi *MLPNN* maka nilai *error* yang dihasilkan akan semkin kecil, sehingga kalimat yang dilafalkan akan semakin bagus kualitas kejelasan dan kenaturalannya.

6. Berdasarkan hasil pengujian dapat dikatakan bahwa sistem *TTS* model prosodi MLPNN memiliki kualitas kejelasan dan pelafalan yang lebih baik dibandingkan dengan sistem IndoTTS.

5.2 Saran

Pada penelitian selanjutnya diharapkan dapat meningkatkan lagi sistem model prosodi dalam menentukan nilai durasi dan pitch fonem. Beberapa hal yang dapat diperhatikan sebagai aspek pengembangan sistem selanjutnya adalah :

1. Pengkajian lebih lanjut pada parameter-parameter dalam sistem *MLPNN* agar informasi yang digunakan sistem untuk proses pembelajaran menjadi lebih lengkap, sehingga dapat meningkatkan kemampuan sistem dalam memprediksi nilai durasi dan pitch.
2. Pengembangan dataset pembelajaran, terutama dari segi variasi fonem yang dijadikan sebagai bahan pembelajaran sistem dan pengkajian ulang terhadap nilai inisialisasi *input* fonem pada dataset pembelajaran secara menyeluruh.
3. Pengembangan mekanisme pembelajaran sistem yang terpisah dalam menentukan nilai durasi dan pitch. Sehingga dapat mempercepat proses pembelajaran. Sebab dalam sistem *TTS* model prosodi *MLPNN* yang sekarang, penambahan dataset berdampak proses pembelajaran sistem menjadi semakin lama.
4. Pengembangan model prosodi yang mampu menangani kalimat yang lebih dari 4 kata dan juga kalimat yang mengandung tanda baca.

DAFTAR PUSTAKA

[1] Agung Nugroho, Fanny, *Tugas Akhir : Perancangan Text to Speech Bahasa Indonesia (Model Prosodi Dataset pada Mbrola),* STT Telkom Bandung, Bandung, 2007.

[2] Arman, Arry Akhmad, IndoTTS, ITB Bandung, Bandung 2000.

[3] Arman, Arry Akhmad, *Teknologi Bahasa dan Perkembangannya di Indonesia. Human Aspects in Computer-Based System Seminars*, Bandung, September 21 - 22, 2005.

[4] Artikel, *http://www.statsoft.com/textbook/stneunet.html/multilayer*, 2003.

[5] Estebon, Michele D., *Perceptrons : An Associative Learning Network*, Virginia tech, Virginia, 1997.

[6] Haidar, Andry, *Thesis : Perancangan dan Implementasi Pemodelan Durasi dan Pitch Berbasis Multilayer Perceptron Neural Network,* Institut Teknologi Bandung, Bandung, 2005.

[7] Kochanski, Greg, Chilin Shih, *Prosody and Prosodic Models*, ICSLP 2002, Denver Colorado, September 16, 2002.

[8] Modul, Artificial Neural Network Exclusive Training 2007, Laboratorium Artificial Intelligence IT Telkom, Bandung, 2007.

[9] Mas'adi, Ali, *Tugas Akhir : Implementasi Perangkat Lunak Server Text to Speech Bahasa Indonesia dengan Unit Ucapan Diphone,* STT Telkom Bandung, Bandung, 2004.

[10] Sami, Lemmetty, "*Reviews of Speech Synthesis Technology*", Helsinki University of Technology, Helsinki, 1999.

[11] T. Dutoit, V. Pagel, N. Pierret, F. Bataille, O. Van derVrecken, *The MBROLA Project : Towards a Set of High Quality Speech Synthesizer Free of Use for Non Commercial Purposes*, InProc, ICSLP 1996, Philadelphia, 1996.

[12] Tebelskis, Joe, *Thesis : Speech Recognation Using Neural Networks*, School of Computer Science, Carnegie Mellon University, Pitssburg, Pennsylvania, 1995.

[13] Vainio, Martti, *Thesis : Artificial Neural Network Based Prosodic Models for Finnish Text-to-Speech Synthesis*, Department of Phonetics, University of Helsinki, Helsinki, 2001.