

PERANCANGAN DAN SIMULASI PENSINTESIS KONKATENASI DAN KONTROL PROSODI BERBASIS JARINGAN SYARAF TIRUAN HOPFIELD

Lintang Mayang Pitaloka¹, Bambang Hidayat², Ratri Dwi Atmaja³

¹Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

Abstrak

Concatenation synthesizer merupakan pensintesis yang mampu memproduksi sinyal ucapan secara otomatis melalui transkripsi grafem-ke-fonem untuk kalimat yang diucapkan. Concatenation synthesizer masih merupakan cara termudah untuk membuat ucapan sistesis berkualitas tinggi.

Penelitian concatenation synthesizer dalam bahasa Indonesia sudah dikembangkan, dan telah mencapai hasil yang cukup memuaskan. Namun demikian beberapa permasalahan masih belum terselesaikan secara tuntas. Oleh karena itu dalam tugas akhir ini dirancang model pensintesis untuk kontrol prosodi untuk meningkatkan kualitas intonasi kalimat pada penelitian sebelumnya. Dalam tugas akhir ini Hopfield Neural Network dipilih untuk merealisasikan model prosodi pada sintesis bahasa Indonesia. Pada bagian pensintesa digunakan model generator intonasi untuk memodifikasi panjang durasi dan kontur intonasi pada setiap suku kata. Digunakan juga Pitch Synchronous Overlap-Add (PSOLA) untuk mengubah pitch yang ada.

Hasil penelitian menunjukkan adanya intonasi kalimat sintesis yang lebih menonjol dibandingkan penelitian sebelumnya. MOS fluidity meningkat dari 2.482353 menjadi 2.81951. MOS intelligibility meningkat dari 2.082353 menjadi 2.77526. Selain itu MOS naturalness juga dapat dipertahankan dari 3.258824 menjadi 3.26143. Akurasi dari pengujian Hopfield Neural Network juga sangat baik, terbukti dengan hasil performansinya yaitu sebesar 100%. Dimana penggunaan. Pada pengujiannya menggunakan FD-PSOLA didapatkan hasil bahwa overlap 90% memiliki pitch countour yang lebih bagus dengan jumlah lebih sedikit sinyal drop pada sambungannya bila di bandingkan dengan overlap yang lain. Untuk TD-PSOLA diperoleh bahwa overlap 1% menjadi titik penyambungan yang paling baik.

Kata Kunci : Hopfield Neural Network, pembangkit prosodi, concatenation synthesizer, pitch synchronous overlap-add.

Abstract

Concatenation synthesizer is a synthesizer that can produce automatic speech signal through the transcription of grapheme-to-phoneme for the spoken sentence. The concatenative synthesis is still become the easiest way the produce a high quality synthetic speech.

Research concatenation synthesizer in the Indonesian version was already developed, and has achieved satisfactory result. However, several problems in the research remain unsolved. In this final task, Hopfield Neural Network (HNN) is selected to realize the concatenation synthesizer model of prosody in the Indonesian language. Pitch Synchronous Overlap-Add (PSOLA) use to modify the signal speech.

The results showed the synthesis of intonation phrase is more prevalent than previous research. Fluidity MOS increased from 2.482353 to 2.81951. Intelligibility MOS increased from 2.082353 to 2.77526. Beside that naturalness MOS also keep from 3.258824 to 3.26143. From the test of Hopfield Neural Network get the best accuracy is 100%. FD-PSOLA test get overlap 90% have more pitch countour and less drop signal at the point of junction if differentiated with other overlap. Overlap 1% to be the best point of junction for TD-PSOLA.

Keywords : Hopfield neural Network, generator prosody, concatenation synthesizer, Pitch Synchronous Overlap-Add.

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, aksent, 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.

Telkom
University

BAB V

PENUTUP

2.1. Kesimpulan

Berdasarkan apa yang telah direalisasikan dan diamati, maka penelitian tentang kontrol prosodi pada pensintesis konkatenasi menggunakan HNN secara umum menghasilkan beberapa kesimpulan sebagai berikut :

1. Sistem sudah berhasil mengucapkan kalimat sintesis yang mempunyai tingkat kejelasan yang lebih baik dari penelitian sebelumnya. MOS *intelligibility* yang diperoleh meningkat dari 2.082353 menjadi 2.77526.
2. Sistem sudah berhasil mengucapkan kalimat sintesis yang mempunyai tingkat kelancaran yang lebih baik dari penelitian sebelumnya. MOS *fluidity* yang diperoleh meningkat dari 2.482353 menjadi 2.81951.
3. Sistem sudah berhasil mengucapkan kalimat sintesis yang mempertahankan intonasi, dimana MOS *naturalness* telah meningkat dari 3.258824 menjadi 3.26143.
4. Pada *level* kalimat, skenario ekstraksi ciri dan HNN sebagai pengontrol intonasi (prosodi) bekerja sangat baik dengan rata-rata akurasi 100% dengan jumlah vektor ciri yang berbeda.
5. Makin panjang vector ciri input maka proses training dan klasifikasi HNN memerlukan waktu komputasi yang lebih lama.
6. Dari ekstraksi ciri yang dirancang menghasilkan rata-rata RMSE sinyal sintesis terhadap sinyal asli untuk 15 kalimat dalam pengujian sebesar 0.20860 dari penelitian sebelumnya sebesar 0.349273. sehingga terjadi peningkatan kemiripan sinyal hasil sintesis dengan sinyal asli.
7. Dari ekstraksi ciri yang dirancang menghasilkan rata-rata koefisien korelasi sinyal sintesis terhadap sinyal asli untuk 15 kalimat dalam pengujian sebesar 0.00392 dari penelitian sebelumnya sebesar 0.004847.

2.2. Saran

1. Pada penelitian selanjutnya disarankan untuk melakukan modifikasi pada domain frekuensi dengan lebih teliti sehingga ucapan kalimat sintesis menjadi lebih dimengerti oleh pendengar, dengan kata lain adalah dengan mengkolaborasi penelitian sebelumnya dengan penelitian kali ini.
2. Pada penelitian selanjutnya dimungkinkan pembuatan algoritma atau teknik baru untuk menentukan nilai parameter dari generator intonasi sehingga lebih memudahkan tanpa harus mencobanya satu persatu sehingga hasilnya juga lebih akurat.
3. Pada penelitian selanjutnya diharapkan dapat memperbaiki modul generator intonasi yang sudah dirancang, khususnya pada titik persambungan antar sinyal sintesis untuk menghindari klik/pop dan suara yang memantul.
4. Perekaman database unit ucapan selayaknya lebih lengkap lagi agar pada saat penggunaan PSOLA lebih mudah dilakukan.
5. Pada penelitian selanjutnya disarankan untuk mencoba menggunakan ciri sinyal suara selain delta magnitude, dengan harapan nilai RMSE dan koefisien korelasi menjadi lebih baik dan bukan tidak mungkin konsumsi memori menjadi lebih hemat.

Telkom
University

DAFTAR PUSTAKA

- [1]. Adipranata, Rudy, dkk. 2006. *Perbandingan Algoritma Exhaustive, Algoritma Genetika Dan Algoritma Jaringan Syaraf Tiruan Hopfield Untuk Pencarian Rute Terpendek*. Universitas Kristen Petra. Surabaya.
- [2]. Anggeraini H, Diah. *Algoritma Jaringan Syaraf Tiruan Hopfield dan Penerapannya pada Travelling Salesman Problem*. Internet.
- [3]. Ananto I.C, Dwi. 2004. *Kontrol Persajakan pada Texto-Speech Bahasa Indonesia Berbasis Recurrent Neural Network*. ITTelkom. Bandung.
- [4]. Departemen Pendidikan dan Kebudayaan. 2001. *Kamus Besar Bahasa Indonesia*. Edisi ketiga. Balai Pustaka. Jakarta.
- [5]. Dwi A, Ratri. 2011. *Perancangan dan Simulasi Kontrol Prosodi pada Concatenation Synthesizer Menggunakan Metode Jaringan Syaraf Tiruan Hopfield*. ITTelkom. Bandung.
- [6]. Ernawati Septima. 2009. *Aplikasi Hopfield Neural Network Untuk Prakiraan Cuaca*. Universitas Diponegoro. Semarang.
- [7]. Heaton Jeff. 2010. *Introduction to Neural Network with Java*. Heaton Research. Amerika.
- [8]. Hebban. 2009. *Introduction to Neural Network*. Hebb's Learning. Amerika
- [9]. Iwut T, Iwan. 2004. *Text-to-Speech Bahasa Indonesia dengan Pembangkitan Prosodi Menggunakan Metoda Multirate Recurent Neural Network*. ITB. Bandung.
- [10]. Iwut T, Iwan. 2006. *Text-to-Speech Bahasa Indonesia Menggunakan Concatenation Synthesizer Berbasis Fonem*. STTTelkom. Bandung.
- [11]. John R. Deller, dkk. 2000. *Discrete Time Processing of Speech Signal*. Macmillan Publishing. Amerika.
- [12]. Jong Jek Siang. 2005. *Jaringan Syaraf Tiruan Pemrograman Menggunakan Matlab*. Penerbit Andi. Yogyakarta.
- [13]. _____. *Speech Production*. Internet.