

PERANCANGAN DAN SIMULASI MODUL EKSTRAKSI CIRI UNTUK KONTROL PROSODI PADA CONCATENATION SYNTHESIZER

Rasif Sahirul Afandi¹, Rita Magdalena², Ratri Dwi Atmaja³

¹Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

Abstrak

Concatenation synthesizer merupakan synthesizer yang mampu memproduksi sinyal ucapan secara otomatis melalui transkripsi grafem-ke-fonem untuk kalimat yang diucapkan. Penelitian concatenation synthesizer dalam berbagai versi bahasa sedang dikembangkan, dan dalam bahasa tertentu telah mencapai hasil yang cukup memuaskan. Namun demikian beberapa permasalahan dalam concatenation synthesizer masih belum terpecahkan secara tuntas. Pendekatan dari setiap permasalahan ini akan selalu mengacu pada tercapainya concatenation synthesizer yang memenuhi kualitas intelligibility (terdengar jelas) dan naturalness (kealamian).

Dalam Tugas akhir ini telah direalisasikan model ekstraksi ciri yang baru sebagai pembanding dari penelitian sebelumnya dan juga telah direalisasikan metode sederhana sebagai pembanding dengan HNN dalam eksekusi data uji. Dengan ekstraksi ciri ini nilai rata-rata RMSE sinyal sintesis terhadap sinyal asli untuk 15 kalimat dalam pengujian meningkat dari 0.349273 menjadi 0.1919, namun untuk rata-rata koefisien korelasi -0.0060 yang artinya sinyal sintesi berkebalikan dengan sinyal asli. Sedangkan untuk metode sederhana lebih cepat 1.63086 dibandingkan menggunakan HNN sebesar 1.97588. Sedangkan pada penilain MOS, untuk integibility meningkat dari 2.666667 menjadi 2,815385, yang artinya kejelasan suara lebih bisa dipahami. Namun untuk fludity dan naturalness justru menurun dibandingkan penelitian sebelumnya.

Perancangan dan simulasi modul ekstraksi ciri dan metode sederhana ini, bisa menjadi dasar dalam meningkatkan kualitas suara sintesis.

Kata Kunci : Hopfield neural network, pembangkit prosodi, concatenation synthesizer

Abstract

Concatenation synthesizer is a synthesizer that is capable of producing a signal through the transcription of speech automatically grapheme-to-phoneme to sentence spoken. Research concatenation synthesizer in various language versions are being developed, and in particular languages has achieved satisfactory results. However, some problems in concatenation synthesizer is still not completely solved. The approach of each of these issues will always refer to the concatenation synthesizer achievement that meets the quality intelligibility (sounds obvious) and naturalness (naturalness).

In this final project, has realized a new model of feature extraction for comparison of previous studies and also have realized a simple method for comparison with the test data HNN in execution. With this feature extraction of the average value of RMSE synthesis signal to the original signal to 15 sentences in the test increased from 0.349273 to 0.1919, but for the average correlation coefficient of -0.0060, which means the signal Sintesi contrary to the original signal. As for the simpler method is faster 1.63086 1.97588 dibandingkan use of HNN. While the MOS in the assessment, to integibility increased from 2.666667 to 2.815385, which means more comprehensible voice clarity. But for fludity and naturalness is decreasing compared to previous studies

Design and simulation modules feature extraction and the simple method, could be the basis dala improve the quality of synthesized speech.

Keywords : Hopfield neural network, generator prosodic, concatenation synthesizer

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^[6]. 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 halusya titik persambungan yang kadang terdengar dalam ucapan sintesis sebagai klik/pop.

Pendahuluan

1.2 Perumusan Masalah

Berdasarkan latar belakang penyusunan Tugas Akhir yang telah diuraikan sebelumnya, permasalahan yang dihadapi dirumuskan sebagai berikut:

1. Bagaimana merealisasikan modul ekstraksi ciri untuk kontrol prosodi pada concatenation synthesizer.
2. Bagaimana merealisasikan HNN dan metode sederhana.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, maka dapat dirumuskan tujuan dari Tugas Akhir kali ini adalah:

1. Merancang modul ekstraksi ciri untuk kontrol prosodi pada concatenation synthesizer.
2. Membandingkan metode HNN dengan metode sederhana.

1.4 Batasan Masalah

Untuk mempermudah dan membatasi cakupan pembahasan masalah pada Tugas Akhir kali ini, maka diberikan batasan-batasan 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.
4. Pemanfaatan kamus fonetik dan database rekaman sinyal ucapan hanya seperlunya saja dengan tetap mendukung tercapainya penyelesaian masalah.
5. Simulasi dibuat dengan menggunakan Matlab.
6. Arti kalimat teks masukan tidak dibahas dan tidak menjadi permasalahan.
7. Jumlah kalimat teks masukan dan jumlah kata dalam tiap kalimat dibatasi maksimal 3 kata, sesuai kebutuhan sistem untuk tetap mendukung tercapainya penyelesaian masalah.
8. Modul algorithm rule based tidak direalisasikan.

Pendahuluan

9. Tanda baca yang digunakan hanya titik (.), tanda tanya (?), tanda seru (!).

1.5 Metodologi Penelitian

Metode yang digunakan dalam penyelesaian tugas akhir ini antar lain :

1. Studi Literatur

Pencarian dan pengumpulan literatur – literatur dan kajian – kajian yang berkaitan dengan masalah – masalah yang ada dalam tugas akhir baik berupa artikel, buku referensi, internet dan sumber – sumber lain.

2. Analisis masalah

Menganalisis semua permasalahan yang ada berdasarkan sumber – sumber dan pengamatan terhadap permasalahan tersebut.

3. Perancangan dan realisasi

Membuat perancangan dan realisasi modul ekstraksi ciri dan metode sederhana.

4. Pengujian dan analisa

Membahas mengenai rincian dari hasil dan analisa modul ekstraksi ciri dan metode sederhana yang direalisasikan.

5. Mengambil kesimpulan

Mengambil kesimpulan dari hasil analisis dan dari masalah yang terjadi.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisi uraian singkat mengenai latar belakang permasalahan, perumusan masalah, maksud dan tujuan penelitian, batasan masalah, metode penelitian serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berupa uraian konsep dan teori dasar secara umum yang mendukung dalam pemecahan masalah, baik yang berhubungan dengan sistem maupun perangkat.

Pendahuluan

BAB III PERANCANGAN DAN REALISASI

Pada bab ini dibahas mengenai perancangan dan realisasi modul ekstraksi ciri dan metode sederhana.

BAB IV PENGUJIAN DAN ANALISA

Bab ini menguraikan pengujian dan analisa prinsip kerja sistem yang telah diimplementasikan. Pengujian dan analisa sistem akan mengacu pada spesifikasi yang telah disebutkan untuk mengetahui apakah hasil rancangan.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dan saran terhadap hasil yang diperoleh dari penelitian yang telah dilakukan.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil perancangan,realisasi dan pengujian sistem, maka dapat disimpulkan bahwa :

1. Metode sederhana mampu mempercepat waktu eksekusi data uji dengan nilai sebesar 1.63086 dibandingkan dengan menggunakan HNN sebesar 1.97588.
2. Modul generator intonasi yang dirancang masih bekerja kurang baik dengan rata-rata RMSE sinyal sintesis terhadap sinyal asli untuk 15 kalimat dalam pengujian sebesar 0.1919 dengan rata-rata koefisien korelasi -0.0060.
3. Penilaian MOS untuk intelligibity meningkat dari 2.67 menjadi 2.815.
4. Metode ekstraksi ciri scenario dua menunjukkan bekerja cukup baik dengan lebih menghemat memori bila dibandingkan scenario satu, namun waktu komputasi menjadi lebih lama.

5.2 Saran

Berdasarkan hasil yang telah dicapai, maka dapat diambil beberapa saran untuk pengembangan lebih lanjut, diantaranya :

1. Pada penelitian selanjutnya disarankan untuk melakukan modifikasi pada domain frekuensi 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 diharapkan dapat memperbaiki modul generator intonasi yang sudah dirancang, khususnya pada titik persambungan antar sinyal sintesis.
3. Perekaman database unit ucapan selayaknya didukung oleh data penelitian yang memadai, seperti data durasi rata-rata suku kata dalam kalimat bahasa Indonesia.
4. Pada penelitian selanjutnya disarankan untuk mencoba menggunakan ciri sinyal suara selain delta magnitude, dengan harapan nilai RMSE dan

Kesimpulan Dan Saran

koefisien korelasi menjadi lebih baik dan bukan tidak mungkin konsumsi memori menjadi lebih hemat.



DAFTAR PUSTAKA

1. Iwut T, Iwan. 2006. *Text-to-Speech Bahasa Indonesia Menggunakan Concatenation Synthesizer Berbasis Fonem*. STTTelkom. Bandung.
2. Laurene Fausett. 1994. *Fundamentals of Neural Networks, Architectures, Algorithms, and Applications*. New Jersey: Prentice Hall, Inc.
3. Maher A Sid-Ahmed. 1995. *Image Processing, Theory, Algorithms, and Architectures*. New York: Mc Graw-Hill, Inc.
4. Diyah Puspitaningrum. 2006. *Pengantar Jaringan Syaraf Tiruan*. Yogyakarta: Penerbit Andi.
5. Jong Jek Siang. 2005. *Jaringan Syaraf Tiruan Pemrograman Menggunakan Matlab*. Yogyakarta: Penerbit Andi.
6. Iwut T, Iwan. 2004. *Text-to-Speech Bahasa Indonesia dengan Pembangkitan Prosodi Menggunakan Metoda Multirate Recurrent Neural Network*. ITB. Bandung.
7. Tjondronegoro, suhartono. 2009. *Statistical Signal Processing*. ITB. Bandung.
8. Dwi A, Ratri. 2011. Perancangan Dan Simulasi Kontrol Prosodi Pada Concetaation Synthesizer Menggunakan Metode Jaringan Saraf Tiruan Hopfield. ITTelkom. Bandung.

Telkom
University