

PERANCANGAN TEXT TO SPEECH BAHASA INDONESIA(MODEL PROSODI DATASET PADA MBROLA)

Fanny Agung Nugroho¹, M.ramdhani², Iwan Iwut Tirtoasmoro³

¹Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

Abstrak

Text to Speech (TTS) adalah suatu sistem pengkonversi teks menjadi suara atau ucapan manusia. Akan tetapi pada teknologi Text to Speech, ada beberapa kendala teknis yang harus diatasi untuk bisa meningkatkan kualitas suara yang dihasilkan, salah satunya adalah dengan penambahan kosakata yang dapat dikenali oleh sistem serta kata atau kalimat yang diucapkan bisa memiliki prosodi atau intonasi. Untuk itu diperlukan cara bagaimana penambahan kosakata dan pengucapan kalimat yang memiliki prosodi pada sistem TTS dapat dilakukan. Sistem Text to Speech secara garis besar terdiri dari dua sub sistem utama, yaitu: NLP (Natural Language Processing) atau Text to Phoneme dan DSP (Digital Signal Processing) atau Phoneme to Speech. Pada tugas akhir ini bagian Text to Phoneme dari sistem TTS ini yang telah dibuat, dimana pada bagian ini dibuat GUI (Grafik User Interface) dari sistem TTS dan algoritma pemrograman untuk : mengenali teks, membuat kamus untuk kata-kata khusus serta bagaimana cara agar prosodi suara dapat dihasilkan. Sedangkan pada bagian Phoneme to Speech, sub sistem TTS ditangani oleh aplikasi MBROLA yang digunakan sebagai speech synthesizer. Pada tugas akhir ini telah direalisasikan Text to Speech bahasa Indonesia dengan model prosodi dataset pada MBROLA, yang dapat mengetahui cara meningkatkan kualitas suara yang dihasilkan untuk pengembangan aplikasi selanjutnya.

Kata Kunci : -

Abstract

Text to Speech (TTS) is a text conversion system to translate text-based material into voice or speech. However, Text to Speech technology present several technical difficulties to overcome in order to provide better voice quality. Among the alternatives are the addition of more computer-recognized words and intonation-imbued sentences. In the end, the question is the matter of how to actually applied those alternatives.

The Text to Speech system is basically divided into two main subsystems: The NLP (Natural Language Processing) or Text to Phoneme and DSP (Digital Signal Processing) or Phoneme to Speech. The aim of this final assignment is to build the Text to Phoneme subsystem which will be equipped with a GUI (Grafik User Interface) and a custom made programming algorithm to recognize text, build custom dictionaries for specific words, and to calculate the production of the actual voice translation. For the Phoneme to Speech subsystem, MBROLA application will be use as a speech synthesizer.

During the research of this final assignment, a working prototype of a Text to Speech system in Indonesian Language was successfully created with MBROLA model dataset, which can also be used to determine a method to enhance voice quality for future researches.

Keywords : -

BAB I

PENDAHULUAN

1.1 Latar Belakang

Aplikasi *Text to Speech* saat ini sudah banyak digunakan dan dikembangkan untuk berbagai macam aplikasi dan kebutuhan serta dengan berbagai macam bahasa yang digunakan termasuk bahasa Indonesia. Namun dalam pengembangan aplikasinya, terutama dalam bahasa Indonesia perlu diperhatikan faktor-faktor yang mempengaruhi kualitas suara, hal ini bertujuan agar suara yang dihasilkan dapat dimengerti oleh manusia dan memiliki kealamian dalam pengucapannya. Ada beberapa cara untuk meningkatkan kualitas suara yang dihasilkan diantaranya adalah penambahan *database* kata dengan harapan teks yang kita tuliskan atau yang kita ketikkan dapat dikenal oleh sistem, serta melakukan perubahan nilai durasi dan *pitch* pada kode fonem sehingga suara yang dihasilkan diharapkan terdengar lebih natural. Cara untuk memperbaiki kualitas suara tersebut dilakukan pada bagian teks ke fonem (*Text to Phoneme*), fonem adalah bagian terkecil dari suatu kata atau ucapan, seperti : /a/, /b/, /c/, /d/, dan seterusnya.

Secara garis besar sistem *Text to Speech* terbagi atas dua blok diagram, pertama adalah blok diagram teks ke fonem (*Text to Phoneme*) dan bagian fonem ke suara (*Phoneme to Speech*). Dan pada tugas akhir ini penulis mengkhususkan diri mengembangkan bagian teks ke fonem dengan harapan kinerja sistem *Text to Speech* dapat ditingkatkan dan sesuai dengan aplikasi yang dibutuhkannya. Sedangkan pada bagian yang lainnya atau bagian *Phoneme to Speech* dilakukan oleh *engine* yang bernama MBROLA yang didalamnya terdapat *diphone database* bahasa Indonesia.

Bagian *Text to Phoneme* pada prinsipnya adalah melakukan konversi dari simbol-simbol tekstual menjadi simbol-simbol fonetik, yang merepresentasikan unit bunyi terkecil (fonem) dari suatu kata, dan simbol-simbol fonetik tersebut dikodekan berdasarkan durasi dan *pitch* yang kita sebut sebagai kode fonem. Dan hanya dengan kode fonemlah *engine* MBROLA dapat mengenal *inputan* teks yang kita ketikkan, yang selanjutnya mengubahnya menjadi suara. Dengan pengembangan bagian *Text to Phoneme* tersebut diharapkan pembuatan sistem *Text to Speech* bahasa Indonesia pada tugas akhir ini sesuai dengan aplikasi yang dibutuhkannya.

1.2 Tujuan dan Manfaat

1. Tujuan dari tugas akhir ini adalah merancang *Text to Speech* bahasa Indonesia model prosodi dataset pada MBROLA.
2. Manfaat dari Tugas Akhir ini adalah agar sistem ini dapat bekerja sesuai dengan aplikasi yang dibutuhkannya, seperti untuk ponsel yang diharapkan dapat membantu para tuna netra untuk dapat membaca teks dan tuna wicara untuk mengucapkan kata atau kalimat.

1.3 Rumusan Masalah

Untuk meningkatkan kualitas suara yang dihasilkan dari *Text to Speech* ada beberapa permasalahan yang harus diperhatikan, diantaranya:

- Bagaimana kosakata dan kata-kata khusus, seperti singkatan dapat dikenali oleh sistem.
- Dapat mengenali berapa simbol atau tanda dalam alphabet.
- Bagaimana suara yang dihasilkan agar dapat memiliki intonasi.

1.4 Batasan Masalah

Pada tugas akhir ini akan dibatasi pada masalah-masalah sebagai berikut :

- Sistem yang digunakan dalam perangkat lunak ini adalah *Text to Speech* dengan id1 sebagai *diphone database* bahasa Indonesia yang dibuat oleh Arry Akhmad Arman.
- Pelafalan teks dan bahasa yang digunakan harus sesuai dengan aturan tata bahasa Indonesia yang disempurnakan (EYD).
- Jenis suara yang dihasilkan hanya satu jenis suara, yaitu suara pria dengan unit ucap *diphone*.
- Untuk *database* kalimat khusus, pengucapan angka dan kata singkatan terbatas digunakan pada sistem ini.
- Menggunakan bahasa pemrograman delphi 7 sebagai *user interface* dari sistem yang dibuat.
- Intonasi atau prosodi yang dihasilkan berdasarkan nilai dari durasi dan *pitch* yang terdapat dalam kode fonem, dalam kasus ini hanya fonem vokal saja

yang akan dicari nilai durasi dan *pitch*nya karena hanya fonem vokal yang mempunyai karakteristik sinyal suara yang dapat diamati bentuk sinyalnya dan dapat berdiri sendiri untuk menghasilkan ucapan.

- Pengucapan kata atau kalimat tidak menggunakan ekspresi marah, sedih, kalimat tanya maupun kalimat perintah.
- Pengembangan sistem dilakukan pada bagian *Text to Phoneme* dan hanya membahas kerja secara umum bagian *Phoneme to Speech*.

1.5 Metodologi

Metode penelitian Tugas Akhir ini meliputi beberapa tahapan. Tahapan tersebut antara lain sebagai berikut:

I. Studi Literatur

Dilakukan dengan mencari bahan-bahan referensi yang dapat mendukung pembuatan sistem tersebut dan aplikasi apa saja yang dibutuhkan didalamnya, serta merumuskan dan mengkaji masalah dengan berbagai referensi tersebut, yang didapat dari internet maupaun dari perpustakaan.

II. Perancangan sistem

Perancangan sistem dilakukan dengan terlebih dahulu mencari kelebihan dan kekurangan apa saja dari sistem yang sudah ada dan bagian mana dari sistem tersebut yang akan dikembangkan. agar dihasilkan spesifikasi sistem yang sesuai.

III. Pembuatan sistem

Pembuatan awal sistem dilakukan dengan terlebih dahulu membuat *database* kosakata, mengubahnya dalam bentuk kode-kode fonem, dan mengubah-ubah kode tersebut agar diperoleh intonasi yang diinginkan. Selanjutnya mengirimkan kode-kode fonem tersebut agar dapat dibaca oleh aplikasi MBROLA.

IV. Analisa sistem

Sistem akan diuji coba sehingga mampu menunjukkan hasil uji yang baik dan membuktikan kehandalan sistem tersebut yang selanjutnya dianalisa kelebihan dan kekurangan dari sistem tersebut.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam Tugas Akhir ini adalah :

- **BAB I PENDAHULUAN**

Pada bab ini dibahas mengenai latar belakang permasalahan, perumusan masalah, tujuan dan manfaat, batasan masalah, metode penelitian serta sistematika penulisan Tugas Akhir.

- **BAB II DASAR TEORI**

Pada bab ini akan dibahas tentang teori-teori dasar yang digunakan dalam penelitian Tugas Akhir ini.

- **BAB III PERANCANGAN DAN REALISASI**

Pada bab ini akan dibahas mengenai perancangan dan realisasi sistem secara keseluruhan.

- **BAB IV HASIL DAN ANALISIS SISTEM**

Pada bab ini menguji hasil dari implementasi sistem dan menganalisis performansi sistem secara keseluruhan terhadap parameter-parameter unjuk kerja sistem.

- **BAB V KESIMPULAN DAN SARAN**

Pada bab ini berisikan kesimpulan hasil penelitian Tugas Akhir ini dan saran untuk pengembangannya.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

1. Sistem TTS yang telah direalisasikan yang dikhususkan pada bagian NLP (*Natural Language Processing*), dapat ditambahkan *database* kata pada sistem tersebut dan dapat menghasilkan ucapan yang memiliki intonasi atau prosodi.
2. Dari hasil pengujian MOS pada *Text-to-Speech* menunjukkan perbedaan frekuensi *diphone database* dengan frekuensi yang dihasilkan untuk nilai prosodi sangat menentukan *intelligibility* dan kenaturalan kata yang diucapkan dengan nilai MOS sebesar (3) dan (3,14).
3. Dari hasil pengujian MOS pada *Text-to-Speech* menunjukkan kesamaan frekuensi *diphone database* dengan frekuensi yang dihasilkan untuk nilai prosodi sangat menentukan *intelligibility* dan kenaturalan kata yang diucapkan dengan nilai MOS sebesar (3,65) dan (3,75).
4. Semakin banyak nilai *pitch* yang digunakan untuk mewakili nilai frekuensi dari suatu fonem maka prosodi dari sinyal suara yang dihasilkan akan semakin baik.
5. Dari hasil pengujian untuk kata atau kalimat dalam aturan biasa (*rule base*), dapat diucapkan dengan baik.
6. Dari hasil perbandingan dengan Indo TTS, menunjukkan untuk kata yang sama dan orang yang sama serta cara pengucapan yang sama maka belum tentu menghasilkan sinyal suara yang sama, baik itu durasi maupun *pitch*nya.

5.2. Saran

1. Untuk mendapatkan sistem TTS yang handal selain hal tersebut diatas, maka pembuatan *diphone database* harus dibuat sendiri agar nilai durasi dan *pitch* yang dihasilkan untuk nilai prosodi yang kita rekam memiliki frekuensi yang sama dengan frekuensi *diphone databasenya*.
2. Bila kita ingin membuat *database* Id1 sendiri maka dibutuhkan alat perekam yang baik, pengucapan yang jelas serta ruangan atau peralatan yang bebas noise sehingga kualitas suara yang dihasilkan atau diucapkan sistem TTS menjadi lebih baik
3. Untuk kata-kata khusus sebaiknya digunakan *database* yang terpisah dari program utamanya, sehingga apabila ingin memasukkan kata-kata baru tidak harus mengubah program utama. Jadi hanya tinggal memasukan kata dan kode fonemnya saja.
4. Untuk pengembangan selanjutnya harus dicari algoritma agar pendeteksian kata pada kata khusus tidak perlu menggunakan spasi untuk akhir kalimat.
5. Untuk pengembangan selanjutnya diupayakan bisa mengucapkan perbedaan antara kata atau kalimat tanya maupun kalimat perintah.

DAFTAR PUSTAKA

- [1] Arman, Arry Akhmad. 2000. "IndoTTS", ITB Bandung.
- [2] Arisman. 2006. "Aplikasi Sms Berbasis *Text to speech* Pada Pocket PC", Tugas Akhir, STT Telkom Bandung.
- [3] Martina, Inge. 2000. "Delphi 5.0", Gramedia Jakarta.
- [4] Mas'adi, Ali. 2004. "Implementasi Perangkat Lunak Server *Text to speech* Bahasa Indonesia Dengan Unit Ucapan Diphone", Tugas Akhir, STT Telkom Bandung.
- [5] P. Taylor. 2007. "Text to Speech Synthesis", University of Cambridge
- [6] Sami, Lemmetty. 1999. "Review of Speech Synthesis Technology", Helsinki University of Technology.
- [7] Somnavilla, Giacomo. 2005. "SINTESI VOCALE CONCATENATIVA PER L'ITALIANO TRAMITE MODELLO SINUSOIDALE", Padova.
- [8] T. Dutoit. 1996. "A Short Introduction to Text-to-Speech Synthesis", Faculte Polytechnique de Mons, TCTS Lab.
- [9] Woolier , Benjamin , *A Text to Speech System for the IBM PC* , Queensland , 1998 .
- [10] <http://indotts.melsa.net.id>.
- [11] <http://tcts.fpms.ac.be/synthesis/mbrola/mbrcopybin.html>.
- [12] <http://project.uet.itgo.com/textto1.htm>.
- [13] http://www.eng.nus.edu.sg/EResnews/0110/rd/rd_4.html