

ANALISIS DAN IMPLEMENTASI PENGENALAN SUARA OTOMATIS UNTUK BAHASA INDONESIA DENGAN MENAMBAHKAN MODEL BAHASA

Agnesia Sidauruk¹, Suyanto², Warih Maharani³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Sistem pengenalan suara otomatis adalah sistem yang memungkinkan komputer untuk dapat mengenali suara yang diucapkan manusia dan menghasilkan respon. Sistem ini membutuhkan model akustik dan model bahasa untuk mendapatkan sistem pengenalan suara yang baik. Salah satu metode dalam membangun model akustik adalah Hidden Markov Model (HMM). HMM dalam model akustik menggambarkan transisi antara state dari sinyal suara yang telah diolah. Hidden Markov Model Toolkit (HTK) merupakan salah satu toolkit yang mampu membangun dan memanipulasi HMM dari model akustik. Selain itu HTK juga mampu menangani dalam pemodelan bahasa dari sistem pengenalan suara. HTK terdiri dari library untuk membangun setiap tool yang ada. User pengguna HTK hanya menggunakan HTK dari sisi tool saja, tidak mengubah library yang ada.

Pada proses pengujian yang dilakukan, didapatkan hasil terbaik dengan menggabungkan model bahasa dan model akustik. Hasil performansi sistem pengenalan suara terbaik yaitu kebenaran kata sebesar 78.89%, akurasi kata 75.72%, dan kebenaran kalimat 45.73%.

Kata Kunci : sistem pengenalan suara otomatis, model akustik, model bahasa, HMM, HTK

Abstract

Automatic speech recognition system is a system that allows computer to be able to recognize the spoken human speech and generates a response. This system requires acoustic model and language model to get a good speech recognition system.

One method of building acoustic models are Hidden Markov Model (HMM). HMM in the acoustic model describes the transition between the state of the speech signal that has been processed. Hidden Markov Model Toolkit (HTK) is a toolkit can build and manipulate the HMM of acoustic models. In addition, HTK also able to handle the modeling language of the speech recognition system. HTK consists of the libraries to build tools. User use HTK in hand tools only, without any change in library.

In the process of testing, get the best results by combining the language model and acoustic models. The result of the best speech recognition system performance are 78.89% word correct , 75.72% word accuracy and 45.73% sentence correct.

Keywords : automatic speech recognition, acoutic model, language model, HMM, HTK

1. Pendahuluan

1.1 Latar belakang

Perkembangan teknologi yang semakin pesat diiringi juga dengan perkembangan kemudahan untuk mendapatkan data. Setiap data mengandung informasi. Data yang ada pada saat ini salah satunya dapat diperoleh melalui suara. Data suara mengandung suatu informasi, oleh karena itu diperlukan adanya pengolahan data yang baik untuk mendapatkan informasi pada data suara secara cepat dan tepat.

Untuk mendapatkan informasi dari data suara dengan jumlah yang kecil mungkin tidak menjadi masalah. Tetapi akan sangat berbeda jika ingin mengingat informasi dari data suara dengan jumlah yang besar. Salah satu cara untuk mendapatkan data informasi suara dan menyimpan informasi tersebut adalah dengan mengolah data suara menjadi teks. Oleh sebab itu dibutuhkan sistem *automatic speech recognition* (ASR) untuk menangani masalah ini.

Pengenalan suara otomatis (*automatic speech recognition*) adalah sistem yang memungkinkan komputer untuk dapat mengenali suara yang diucapkan manusia dan menghasilkan respon. Penelitian mengenai pengenalan suara otomatis ini sudah dikembangkan diberbagai negara. Tentu saja pengenalan suara otomatis ini dikembangkan sesuai dengan bahasa yang digunakan di negara tersebut. Berdasarkan [6], sistem pengenal suara otomatis merupakan suatu sistem yang bersifat '*language dependent*' artinya suatu sistem atau teknik yang berlaku untuk suatu bahasa dan tidak mudah diterapkan untuk bahasa lainnya, hal ini dikarenakan setiap bahasa memiliki karakteristik yang berbeda-beda. Salah satu karakteristik paling mendasar adalah fonem yang berbeda antara bahasa yang satu dengan bahasa yang lain. Penelitian mengenai pengenalan suara berbahasa Indonesia masih sangat minim. Salah satu penelitian awal sudah dilakukan oleh [8].

Sistem pengenalan suara otomatis membutuhkan model akustik dan model bahasa [8]. Model akustik berhubungan dengan ekstraksi awal suara ketika diinputkan. Suara memiliki karakteristik khusus yang harus diperhatikan, antara lain sinyal pengucapan vokal dan fonem. Model akustik sangat mempengaruhi kualitas sistem pengenalan suara otomatis. Model akustik yang dibangun dengan baik akan menghasilkan sistem pengenal suara otomatis yang baik pula.

Hidden markov model (HMM) merupakan salah satu bentuk pemodelan statistik dalam menganalisa karakteristik sinyal ucapan. Model ini merepresentasikan sinyal ucapan menjadi *state diagram* yang memiliki probabilitas tertentu setiap perubahan *statenya*. Untuk membangun HMM, telah tersedia perangkat lunak *open source* yang dapat digunakan yaitu *Hidden markov model Toolkit* (HTK). Perangkat lunak ini memberi banyak kemudahan pada setiap *user* yang menggunakannya. Pada HTK sudah terdapat modul-modul yang mendukung sistem pengenalan suara baik pengenalan kata maupun fonem.

Untuk menambah performansi dari sistem pengenal suara otomatis tidak hanya membutuhkan pemodelan akustik yang baik namun juga diperlukan model bahasa yang baik. Menurut [4], akurasi pengenalan akan berdampak besar dengan

memanfaatkan rangkaian informasi yang dikenali. Rangkaian informasi ini dapat berupa pengenalan kata yang diterima oleh sistem. Rangkaian informasi pengenalan kata ini dapat diperoleh melalui model bahasa. Model bahasa digunakan dalam *speech recognition* untuk membantu menentukan probabilitas dari urutan hipotesis kata dan akan membuat sistem membatasi ruang pencarian kata ke arah pengenalan urutan kata yang memiliki kemungkinan untuk benar. Salah satu pemodelan bahasa secara statistik adalah N-gram *language model*. N-gram *language model* merupakan model bahasa dimana setiap kata akan dilihat probabilitasnya berdasarkan kemunculan kata tersebut bersama-sama. Pada tugas akhir ini, akan mencoba menerapkan dan menganalisis sistem pengenalan suara untuk bahasa Indonesia dengan menambahkan model bahasa.

1.2 Perumusan masalah

Berdasarkan latar belakang tersebut, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana membangun sistem pengenalan suara otomatis untuk bahasa Indonesia dengan menggunakan HTK.
2. Bagaimana tingkat akurasi dari sistem pengenalan suara otomatis yang dibuat.
3. Faktor-faktor apa saja yang mempengaruhi tingkat keberhasilan akurasi sistem tersebut (membandingkan faktor pemodelan akustik dan pemodelan bahasa).

1.3 Tujuan

Tujuan penyusunan tugas akhir ini adalah :

1. Mengimplementasikan pengenalan suara pada bahasa Indonesia menggunakan HTK.
2. Melakukan pengujian untuk mengukur tingkat akurasi dari sistem pengenalan suara otomatis.
3. Menganalisis faktor-faktor yang mempengaruhi tingkat keberhasilan akurasi sistem dalam hal ini membandingkan antara faktor pemodelan akustik dan pemodelan bahasa.

1.4 Batasan Masalah

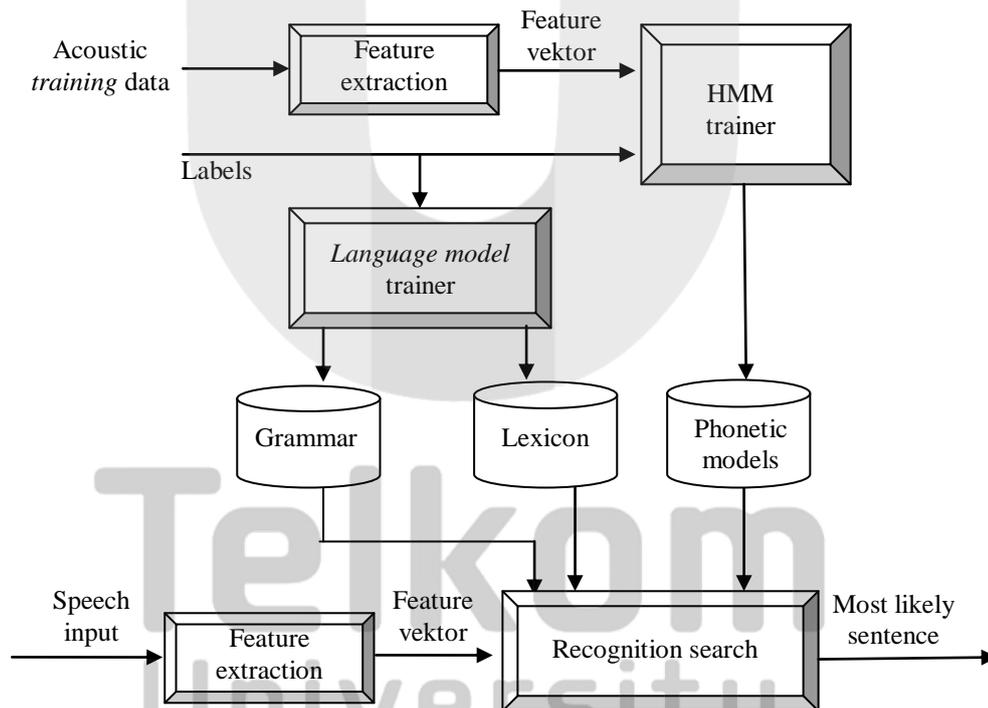
Batasan masalah dalam penelitian tugas akhir ini, antara lain :

1. Tidak melakukan proses perekaman suara. Data suara berekstensi .wav didapat dari TELKOM Risti.
2. Tidak membangun sistem pengenalan suara sendiri melainkan menggunakan HTK berbasis PERL.
3. Sistem pengenalan suara yang dibuat menggunakan transkripsi *monophone* dan *triphone*.
4. Model bahasa yang akan digunakan adalah model bahasa *bigram* dan *trigram*.
5. Tidak memperhitungkan *noise* baik pada saat *training* maupun *testing*.

1.5 Metodologi penyelesaian masalah

Metodologi yang digunakan dalam memecahkan masalah di atas adalah dengan langkah-langkah berikut:

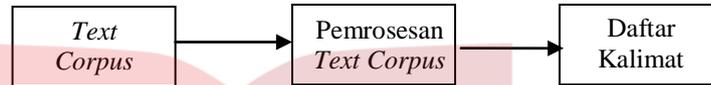
1. Studi literatur
Pencarian referensi dan sumber-sumber yang berhubungan dengan *speech recognition*, HMM, HTK, dan N-gram model dalam menyelesaikan tugas akhir ini.
2. Pengumpulan data
Mengumpulkan data suara dan data teks. Data suara berasal Telkom
3. Analisis dan perancangan sistem
Melakukan analisis dan perancangan terhadap sistem yang dibangun, menganalisa metode yang akan digunakan untuk menyelesaikan permasalahan, termasuk menentukan bahasa pemrograman yang digunakan, arsitektur, fungsionalitas, dan antarmuka sistem. Input sistem adalah rekaman ucapan suara bahasa Indonesia. Output dari sistem adalah *text* hasil ucapan dan presentase akurasi kesesuaian ucapan dengan *text*.
4. Implementasi dan pembangunan sistem
Tahapan dalam pembangunan sistem seperti pada gambar 1-1:



Gambar 1-1 Model bahasa dalam pengenalan suara otomatis [4]

- a. Menyiapkan *speech corpus* yang akan digunakan untuk *training*.
- b. *Speech corpus* yang ada sebelum dibangun model HMM akan dilakukan ekstraksi ciri dengan menggunakan metode MFCC.
- c. *Speech corpus* yang ada diberi label berbasis fonem yang akan digunakan untuk *language model trainer* dan *training*.

- d. Selanjutnya akan dilakukan pembangunan model HMM dan kemudian dilakukan proses pelatihan terhadap *speech corpus*.
- e. Tahap selanjutnya adalah menyiapkan *text corpus* dan melakukan *text preprocessing* (proses mengubah huruf besar menjadi kecil dan pengubahan tanda baca) dan *parsing* (penghilangan tag-tag yang ada) pada *text corpus*. Pemrosesan *text corpus* dapat dilihat pada gambar 1-2.



Gambar 1-2 Proses pembuatan daftar kalimat

- f. Dari *text preprocessing* dibangun model bahasa dan kamus.



Gambar 1-3 Proses pembuatan kamus

- g. Dan tahap terakhir adalah pengujian dengan memasukkan data *speech corpus* untuk *testing* dengan menambahkan kamus dan model bahasa.
5. Analisis
Analisa hasil penelitian yaitu untuk mengetahui akurasi kesesuaian antara kata-kata ucapan yang dimasukkan dengan *text* yang dihasilkan.
 6. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.

5. Kesimpulan

5.1 Kesimpulan

Dari hasil pengujian dan analisis yang dilakukan pada sistem, didapat kesimpulan sebagai berikut:

1. Model HMM yang dibangun dengan *manual labelling* lebih baik dari model HMM yang dibangun dengan *flat start*. Hasil pengenalan kata *manual labelling* lebih besar atau sama dengan *flat start*. Hasil akurasi pengenalan kalimat *manual labelling* untuk pengenalan kata lebih baik dibandingkan dengan *flat start*. Hasil pengenalan kalimat *manual labelling* lebih baik dibandingkan hasil pengenalan kalimat *flat start*.
2. Berdasarkan hasil uji dari perbandingan *flat start* dan *manual labelling* didapat hasil pengenalan kata, pengenalan kalimat, dan akurasi kata tertinggi pada metode *manual labelling triphone*. Untuk metode *manual labelling* dan *flat start triphone* memiliki performa dan akurasi yang lebih baik dibanding *monophone*.
3. Model HMM dengan *language model trigram* mempunyai performa lebih baik dalam pengenalan kalimat dibandingkan dengan *language model bigram*.
4. Semakin besar jumlah korpus teks yang *training* dalam *language model* meningkatkan performa sistem baik dalam pengenalan kata ataupun kalimat.

5.2 Saran

Untuk pengembangan sistem ASR selanjutnya perlu memperhatikan hal-hal berikut ini:

1. Implementasi sistem ASR dengan metode *manual labelling* dengan jumlah data *training* yang besar dan memiliki variasi dialek dari berbagai daerah di Indonesia.
2. Perlu dibuat sistem yang mampu melabeli secara otomatis dengan tingkat keakuratan yang baik.
3. Perlu dibangun sistem yang memiliki korpus teks dengan jumlah besar untuk dilatih pada *language model*.

Daftar Pustaka

- [1] Arman, A.A, *Proses Pembentukan dan Karakteristik Sinyal Ucapan..* Diunduh pada: <http://indotts.melsa.net.id> , 30 September 2010
- [2] Arman, A.A, *Teknologi Pemrosesan Bahasa Alami sebagai Teknologi Kunci untuk Meningkatkan Cara Interaksi antara Manusia dengan Mesin.* Diunduh pada: <http://indotts.melsa.net.id> , 30 September 2010
- [3] Aswin Juari, Ayu Purwarianti. *Deteksi OOV Menggunakan Hasil Pengenalan Suara Otomatis Untuk Bahasa Indonesia.* Diunduh pada: <http://jiki.cs.ui.ac.id> , 10 September 2010
- [4] Becchetti C., Ricotti L.P. *Speech recognition Theory and C++ Implementation.* John Wiley & Sons. 1999
- [5] Digalakis, V., Oikonomidis, D., Pratsolis D., Tsourakis, N., Vosnidis, C., Chatzichrisafis, N., and Diakouloukas, V. *Large Vocabulary Continuous Speech recognition in Greek: Corpus and Automatic Dictation System.* Diunduh pada: 10 September 2010
- [6] Fiarni, Cut. *Analisis Efektivitas Pengembangan Korpus Speech recognition Bahasa Indonesia Menggunakan Pendekatan Iteratif.* Thesis EL ITB. 2007. Diunduh pada: <http://digilib.itb.ac.id> , 10 September 2010
- [7] Maharani, Warih. *Analisis Performansi Recognition Experimental System (RES) untuk Bahasa Indonesia.* Diunduh pada: <http://yudiagusta.files.wordpress.com> , 30 September 2010
- [8] Puji Lestari, Dessi, Koji Iwano, Sadaoki Furui. *A Large Vocabulary Continuous Speech recognition System For Indonesian Language.* Diunduh pada: www.furui.cs.titech.ac.jp/publication/2006/ISA-Japan2006-17.pdf , 10 September 2010
- [9] Rabiner, Lawrence, Biing-Hwang Juang. *Fundamental of Speech recognition.* Prentice Hall, 1993
- [10] Syafur, Irfan. *Aplikasi Pengenalan Wicara HMM untuk Kendali Robot PDA.* Proyek Akhir Teknologi Komunikasi Politeknik Elektronika Negeri Surabaya. 2007
- [11] Tunali, Volkan. *A S peaker Dependent, Large Vocabulary, Isolated Word Speech recognition System for Turkish.* Thesis Computer Engineering Marmara University. 2005. Diunduh pada: www.volkantunali.com/dosyalar/VOLKAN_TUNALI_tez.pdf , 30 September 2010
- [12] Young S.J et.al. , *The HTK Book (for HTK version 3.4)*, Diunduh pada: <http://eng.cam.ac.uk>