

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

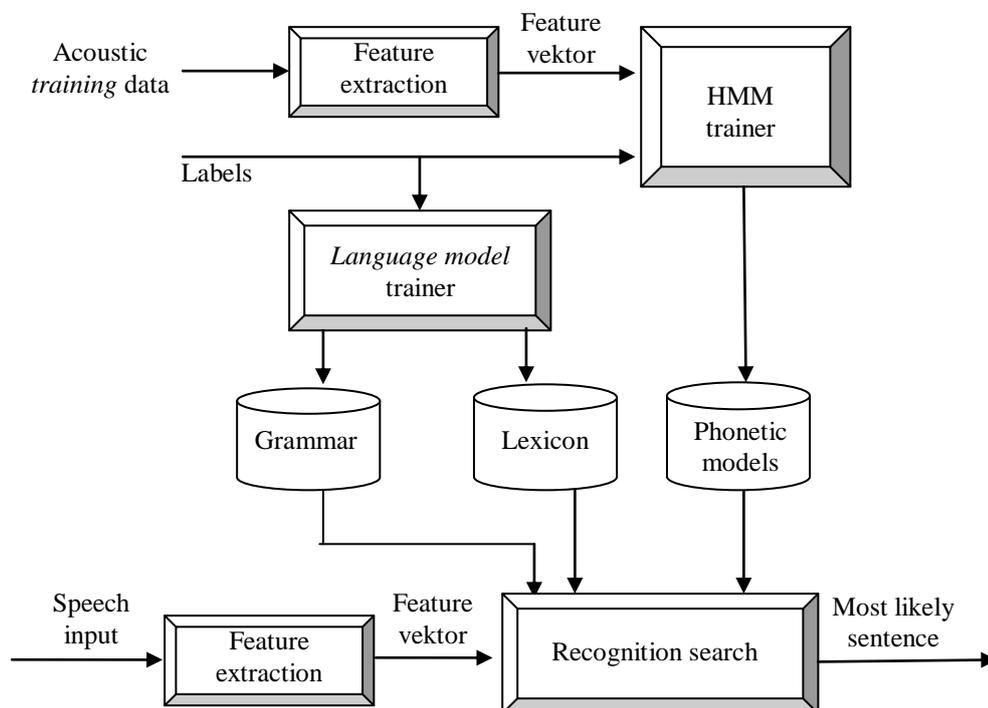
Batasan masalah dalam penelitian tugas akhir ini, antara lain :

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:

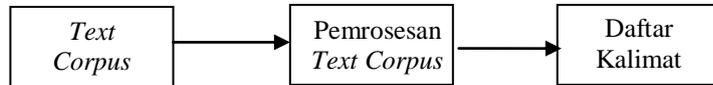
1. Studi literatur
Pencarian referensi dan sumber-sumber yang berhubungan dengan *speech recognition*, HMM, HTK, dan N-gram model dalam menyelesaikan tugas akhir ini.
2. Pengumpulan data
Mengumpulkan data suara dan data teks. Data suara berasal Telkom
3. Analisis dan perancangan sistem
Melakukan analisis 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.