

## **Analisis dan Perancangan *Speech Recognition* Translate Bahasa Inggris-Indonesia-Gorontalo Berbasis Android**

**Ismail Mohidin**

Program Studi Teknik Informatika, Politeknik Gorontalo  
Jl. JL. Muchlis Rahim, Desa Panggulo, Kec. Botupingge, Kab. Bone Bolango 60111  
Email : [is.mohidin@poligon.ac.id](mailto:is.mohidin@poligon.ac.id)

Terima Naskah : 5 April 2016

Terima Revisi : 29 April 2016

### **ABSTRAK**

Aplikasi *Translate* merupakan sebuah rancang bangun aplikasi *speech recognition* yang dibangun pada sistem operasi Android. Aplikasi ini dirancang menggunakan sebuah *smartphone* sebagai perangkat keras untuk menerima input berupa kata yang diucapkan oleh *user*. Metode yang digunakan adalah *Hidden Markov Model* (HMM) yang terdapat pada *library Pocketsphinx*. Fonem yang digunakan adalah aturan fonem Indonesia. Kelebihan dari aplikasi ini adalah dapat digunakan tanpa akses internet.

Penulis mencoba membangun aplikasi *Translate* bahasa Inggris-Indonesia-Gorontalo menggunakan *speech to text* berbasis android yang akan mengenalkan kata yang akan diucapkan oleh *user* dalam tiga bahasa dan diikuti pengucapan, Aplikasi ini diharapkan dapat membantu dan menunjang perkembangan budaya dan bahasa daerah Gorontalo.

*Kata kunci: Speech Recognition, Translate Bahasa, Hidden Markov Model, Pocketsphinx, Android*

### **ABSTRACT**

*Translate app is a plan to build speech recognition applications built on the Android operating system. This application is designed to use a smartphone as a hardware to receive input in the form of words spoken by the user. The method used is the Hidden Markov Model (HMM) contained in Pocketsphinx library. Phoneme used are phonemes Indonesian rule. The advantages of this application is that it can be used without internet access.*

*The author tries to build applications Translate English Indonesian-Gorontalo using speech to text-based android that will introduce the word to be spoken by the user in three languages and followed the pronunciation, this application is expected to assist and support the development of local culture and language Gorontalo.*

*Keywords: Speech Recognition, Translate, Hidden Markov Model, Pocketsphinx, Android*

### **PENDAHULUAN**

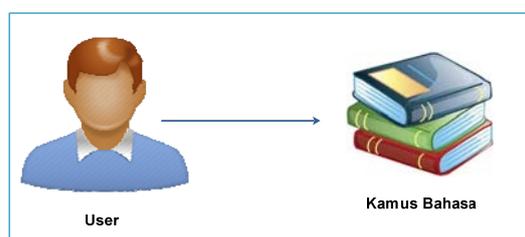
Seiring berkembangnya waktu, ponsel cerdas (*smart phone*) sudah bukan lagi barang baru bagi masyarakat saat ini karena semua aktifitas manusia tidak lepas dari penggunaan ponsel cerdas (*smart phone*). varian yang paling banyak diminati, hal ini dikarenakan banyak aplikasi populer dan banyak digunakan salah satu contoh aplikasi yang populer digunakan adalah aplikasi penerjemah bahasa.

Provinsi Gorontalo adalah provinsi yang memiliki beragam bahasa yang digunakan sebagai contoh bahasa bone, tilamuta, suwawa. Meskipun

memiliki beragam bahasa, masyarakat gorontalo umumnya menggunakan bahasa asli Gorontalo.

Dengan beragam daerah dan bahasa terdapat beberapa kendala saat berkomunikasi salah satunya adalah perbedaan bahasa. Penggunaan buku kamus bahasa masih menjadi salah satu pilihan dalam menerjemahkan bahasa. Namun dalam penggunaannya sebagian masyarakat memiliki kendala yaitu ketika akan menerjemahkan bahasa asing yang sulit untuk ditulis dan diucapkan seperti bahasa Gorontalo, Inggris dan Indonesia dan lain sebagainya akan mempersulit masyarakat dalam menerjemahkan bahasa. Selain itu Penggunaan

buku kamus bahasa masih menjadi salah satu media dalam menangani perbedaan bahasa. Dalam proses penggunaannya, pengguna diharuskan mencari satu persatu kata/kalimat dari sebuah percakapan yang hendak di terjemahkan yang pada umumnya tersusun berdasarkan abjad. Prosedur ini membutuhkan waktu yang relatif lama dikarenakan dalam pencariannya dilakukan secara manual. Selain itu, pengguna akan disulitkan dengan lawan bicara yang memiliki bahasa yang sulit untuk ditulis dan diucapkan seperti bahasa china, korea, jepang, rusia dan lain sebagainya.



Gambar 1. Alur Penggunaan Buku Kamus Bahasa

Penggunaan situs penterjemah bahasa seperti Google Translator ataupun Bing Translator yang tersedia pada <http://translate.google.com/> dan <http://www.bing.com/translator> sangat familiar dalam menerjemahkan bahasa. Dalam penggunaannya, cara/prosedur ini relatif lebih cepat dalam menerjemahkan bahasa dibanding menggunakan buku kamus bahasa. Untuk dapat menerjemahkan bahasa menggunakan *Google Translator* ataupun *Bing Translator*, pengguna diharuskan memiliki koneksi internet dalam pengoperasiannya.



Gambar 2. Alur Penggunaan Situs Penterjemah Bahasa

Untuk mengatasi kendala-kendala tersebut, pertama dibutuhkan sebuah aplikasi penterjemah bahasa yang menerapkan teknologi *speech to text* yang dapat mengkonversi ucapan menjadi teks membantu masyarakat dalam menerjemahkan

bahasa yang sulit untuk ditulis dan diucapkan. Dengan memanfaatkan *Pocketsphinx* merupakan *library* pengenalan ucapan versi mobile application dari sistem *Sphinx*. aplikasi yang akan dibangun diharapkan dapat menangani kendala tersebut, proses penterjemahan dapat dilakukan melalui suara (ucapan). Kedua, dibutuhkan suatu aplikasi yang dapat digunakan dalam kondisi terhubung internet (*online*) maupun tidak terhubung internet (*offline*). Dalam hal ini, untuk penggunaan secara *offline* pada aplikasi akan dibangun sebuah fitur *Phrasebook* yang berisi files kata umum yang sering digunakan pada saat berkomunikasi yang terdiri dari beberapa kategori. Ketiga, dengan dibangunnya sebuah aplikasi berbasis mobile, maka aplikasi dapat digunakan kapanpun dan dimanapun. Berdasarkan uraian yang dikemukakan, maka terdapat masalah yaitu bagaimana menerapkan teknologi *speech to text* pada aplikasi penterjemah bahasa Gorontalo, Inggris dan Indonesia dengan memanfaatkan *pocketsphinx* guna membantu dalam menerjemahkan bahasa yang sulit untuk ditulis dan diucapkan, bagaimana membuat fitur *Phrasebook* yang berisi *files* kata umum yang sering digunakan pada saat berkomunikasi yang terdiri dari beberapa kategori yang dapat digunakan secara *offline*, serta bagaimana membuat sebuah aplikasi berbasis mobile yang dapat diakses. Data yang akan digunakan dalam dalam penelitian ini berupa data suara berbahasa Indonesia, Inggris dan Gorontalo sebanyak 30 data suara.

### *Speech To Text*

Menurut Taylor dan Paul (2007) *Speech To Text* Merupakan salah satu tipe dari aplikasi *speech synthesis* digunakan untuk membuat versi suara dari teks mencakup juga *voice-enabled-email* dan perangkat suara pada sistem dengan respons suara. *Speech to text* juga sering kali digunakan bersamaan dengan program *voice recognition*. *Voice Recognition* adalah suatu sistem yang dapat mengidentifikasi seseorang melalui suaranya, *Voice Recognition* tidak mengidentifikasi siapa yang akan berbicara, tetapi *speech recognition* mengidentifikasi apa yang diucapkan. Prinsip kerja aplikasi *speech to text* menerima gelombang suara manusia lalu mengidentifikasikan setiap karakter pada katanya lalu mengkonvertnya menjadi sebuah text : contoh aplikasi sederhananya adalah aplikasi *Speech*

*Recognition* yang bisa mengidentifikasi suara kita berdasarkan tiap karakternya menjadi text.

### **Pocketsphinx**

Pocketsphinx merupakan *library* pengenalan ucapan versi *mobile application* dari sistem *Sphinx* yang dirancang oleh *Carnegie Mellon University*. Metode yang digunakan dalam sistem *speech recognition Pocketsphinx* ini yaitu metode *Hidden Markov Model*. Proses pembelajaran unit-unit suara disebut *training*, sedangkan proses menggunakan pengetahuan yang diperoleh untuk menyimpulkan urutan yang paling mungkin dari unit dalam sinyal yang diberikan disebut *decoding*, atau secara sederhana disebut pengenalan (*recognition*). Karena terdapat dua proses tersebut maka diperlukan *SPHINX trainer* dan *SPHINX decoder*.

### **Android**

Android adalah system operasi berbasis linux. Android memiliki keunggulan sebagai perangkat lunak yang dapat didistribusikan secara terbuka (*open source*) sehingga pengguna bisa membuat aplikasi baru di dalamnya.

Amrin Hakim, Muhammad. (2011) Fitur yang tersedia di Android adalah:

- Kerangka aplikasi: itu memungkinkan penggunaan dan penghapusan komponen yang tersedia.
- Dalvik mesin virtual: mesin virtual dioptimalkan untuk perangkat mobile.
- Grafik: grafik di 2D dan grafis 3D berdasarkan pustaka OpenGL.
- SQLite: untuk penyimpanan data.

### **METODE**

Sistem yang digunakan pada aplikasi ini menitik beratkan pada pengguna. Pengguna harus memasukkan data suara (untuk proses *speech to text*). Teks yang dimasukkan juga harus sesuai dengan bahasa yang akan digunakan, karena teks yang dimasukkan *case sensitive*. Bahasa atau audio yang dihasilkan sesuai dengan bahasa yang diinputkan. Kemudian untuk suara yang dimasukkan juga harus sesuai dengan bahasa yang dipilih.

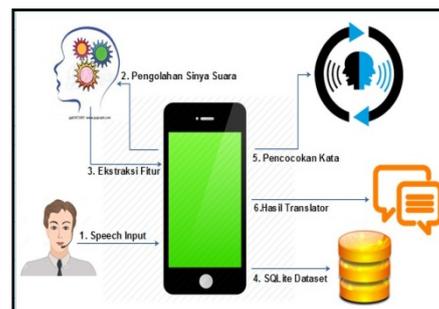
Saat ini tersedia aplikasi mobile yang menyediakan fasilitas penterjemahan bahasa. Dalam penggunaannya, cara/prosedur ini lebih praktis karena pengguna tidak terlebih dahulu harus membuka aplikasi browser untuk dapat menerjemahkan bahasa.



Gambar 3. Alur Penggunaan Aplikasi Penterjemahan Bahasa

### **Analisis Sistem Yang Dibangun**

Sistem yang akan dibangun adalah aplikasi translate berbasis mobile yang dapat dioperasikan menggunakan platform Android. Dalam pembangunannya, aplikasi ini mengimplementasi teknologi *speech to text* yang bertujuan untuk mempermudah pengguna dalam berkomunikasi dengan lawan bicara yang memiliki bahasa yang sulit untuk ditulis dan diucapkan, seperti bahasa Inggris, Indonesia dan Gorontalo. Untuk menunjang kegiatan yang memiliki mobilitas tinggi, aplikasi yang akan dibangun dapat digunakan dalam kondisi online maupun offline. Dalam penggunaan online, Bahasa dan hasil terjemahan bahasa dapat disimpan kedalam fitur *Favourites* yang nantinya dapat diakses secara *offline*. pengguna dapat menerjemahkan percakapan melalui *Speech Recognition*. Gambar 4. menunjukkan gambaran umum sistem yang akan dibangun.



Gambar 4. Alur Kerja Sistem Yang Akan Dibangun

Dalam analisis pengolahan sinyal suara dibagi menjadi beberapa tahap. Tahap pertama adalah pemilteran sinyal suara yang berupa sinyal analog dan mengubah sinyal analog ke digital, yaitu dengan proses konversi analog ke digital (ADC). Tahap kedua adalah tahap mengekstrak file suara yang menghasilkan sebuah informasi

dari file suara tersebut dengan *Fast Fourier Transform* (FFT). Tahap ketiga adalah mencocokkan file suara tersebut dengan sample yang sesuai menggunakan *Hidden Markov Models* (HMM).

**Konversi Analog ke Digital (ADC)**

Sinyal suara yang akan diproses bersifat analog sehingga jika dilakukan pengolahan secara digital, sinyal suara tersebut harus dikonversi menjadi sinyal digital, berupa urutan angka dengan tingkat presisi tertentu yang dinamakan analog to digital conversion dengan menggunakan analog-to-digital converter (ADC). Konsep kerja ADC terdiri dari tiga proses, yaitu:

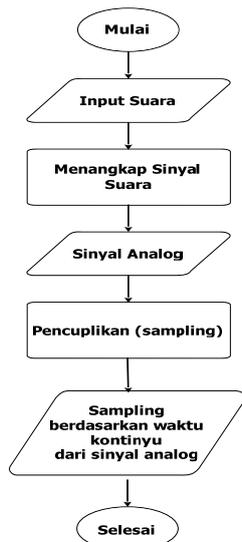


Gambar 5. Konsep Kerja ADC (Analog to Digital Converter)

Keterangan konsep kerja ADC :

1. Sampling

Pada tahap ini yang pertama dilakukan ialah mengambil sample yaitu dengan amplitudo yang mengarah ke atas dari sumbu x, lalu mengambil nilai titik-titik diskrit dari sinyal waktu kontinu supaya mudah untuk dikuantisasi. Titik diskrit merupakan data yang satuannya selalu bulat dalam bilangan asli, tidak berbentuk pecahan. Proses sampling ini mempunyai alur yang<sup>3</sup> digambarkan pada gambar 6 berikut.

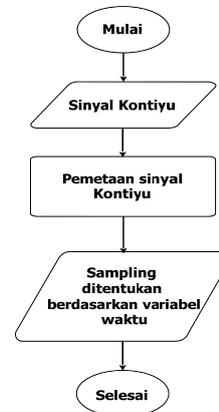


Gambar 6. Flowchart sampling

Dalam proses sampling ini dilakukan pada perintah suara sebanyak 30 kata antara lain untuk pengucapan bahasa Indonesia, Inggris dan Gorontalo dengan panjang durasi rekaman selama 3 detik. Data berupa sinyal suara diperoleh dengan cara merekam suara melalui *microfon smartphone* Android. Suara tersebut mempunyai format .wav dengan frekuensi sampling adalah 800Hz.

2. Kuantisasi

Kuantisasi merupakan proses pemetaan dari nilai sinyal kontinu menjadi nilai-nilai yang diskrit sehingga didapatkan sinyal nilai diskrit, dan sinyal diskrit diambil dari hasil sampling.



Gambar 7. Flowchart kuantisasi

**Coding**

Pada proses ini, tiap nilai diskrit yang telah didapat, dipresentasikan dengan angka biner. Gambar 8. di bawah ini merupakan alur proses coding.



Gambar 8 Flowchart coding

**HASIL DAN PEMBAHASAN**

**Analisis Ekstraksi Informasi**

Analisis sinyal adalah kegiatan melakukan ekstraksi terhadap semua informasi yang terdapat di suatu sinyal. Adapun langkah-langkah analisis LPC untuk pengenalan suara adalah sebagai berikut:



Gambar 9. Diagram blok langkah-langkah LPC

1. *Preemphasis* (penekanan sinyal) adalah suatu proses produksi suara manusia, radiasi pada bibir dan lidah ketika proses phonation mengakibatkan komponen frekuensi tingginya. Tujuan *preemphasis* adalah untuk memperbaiki sinyal dari gangguan *noise* sehingga meningkatkan akurasi pengenalan suara untuk tahap selanjutnya.

2. *Framming*  
 Pada tahap ini sinyal hasil *preemphasis* dikelompokkan ke dalam bingkai dengan ukuran masing-masing bingkai sebesar N data. Bingkai ini berurutan dengan pemisahan antara kedua bingkai sebesar M data. Biasanya  $M = 1/3 N$ .

3. *Windowing*  
 Melakukan *windowing* terhadap setiap frame yang telah dibentuk untuk meminimalkan diskontinuitas pada ujung awal dan ujung akhir setiap frame. Proses *Fast Fourier Transform* (FFT)

*Fast Fourier Transform* (FFT) tujuan adalah mengekstrak file suara yang menghasilkan sebuah informasi dari file suara tersebut. *Fast Fourier Transform* (FFT) merupakan algoritma yang mengimplementasikan *Discreet Fourier Transform* (DFT) dengan teknik perhitungan yang cepat serta memanfaatkan sifat *periodikal* dari *transformasi fourier*.

**Analisis SQLite Database**

Pada fitur ini memungkinkan pengguna untuk mengakses kembali terjemahan percakapan yang telah dilakukan sebelumnya pada fitur *Translator* secara *offline*. Output pada fitur *Translator* berupa percakapan awal, percakapan hasil terjemahan, dan suara hasil terjemahan akan menjadi input pada fitur *Favourites*. Pesan percakapan tersebut akan di-input-kan ke dalam

SQLite Database *Favourites* yang nantinya akan digunakan pada fitur *Favourites*.

1. *Save Conversation*

Ketika pengguna menjalankan fungsi *Add To Favourites* terdapat 5 parameter yang disimpan kedalam database yaitu bahasa asal, bahasa tujuan, teks percakapan asal, teks percakapan hasil terjemahan, dan alamat/*path* audio dari percakapan hasil terjemahan. sistem akan membuka koneksi database, lalu memproses *query insert*, setelah proses *insert* selesai kemudian sistem akan menutup koneksi database.

2. *Request*

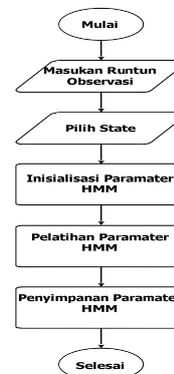
Ketika pengguna menjalankan fungsi *Favourites* sistem akan membaca seluruh data yang berada pada database.

3. *Result*

Sistem akan merespon dengan memberikan data sesuai apa yang telah *direquest* sebelumnya dalam bentuk *ArrayList* dan menampilkannya dalam bentuk *ListView*. Data tersebut meliputi teks percakapan asal, bahasa asal, teks percakapan hasil terjemahan, dan bahasa tujuan. Ketika pengguna memilih salah satu *ListView* sistem akan membaca alamat/*path* audio yang berada pada database yang kemudian menjalankan file audio berformat *.wav* (percakapan hasil terjemahan).

**Analisis Pencocokan Kata**

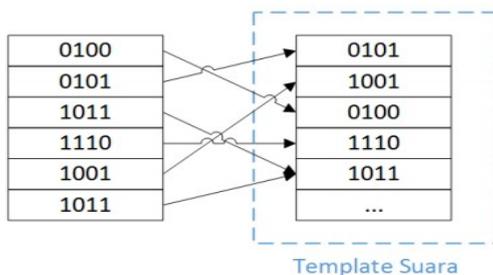
Dalam pencocokan kata menggunakan metode *Hidden Markov Models* (HMM), dimana HMM ini bertugas untuk memastikan pencocokan file suara dengan *template* suara yang tersedia. Ada beberapa tahap yaitu ekstraksi sinyal suara dengan menggunakan FFT dan menentukan nilai probabilitas yang maksimum dengan menggunakan *Hidden Markov Models* (HMM). Pencocokan data suara baru dan hasil ekstraksi akan menggunakan teknik *binary search*.



Gambar 10. Alur Proses Pencocokan Kata

Proses pengenalan ucapan (*recognition*) dilakukan pada HMM. Dalam pemrosesannya untuk dapat mengkonversi suara ke dalam bentuk teks terdapat beberapa tahapan, diantaranya:

1. Suara pengguna berupa sinyal analog dikonversi oleh *device* kedalam bentuk diskrit yang kemudian di ubah bentuk kedalam bentuk biner dan secara *realtime* dikirim ke HMM untuk dilakukan proses konversi
2. Setelah pengguna berhenti berbicara (*stop recording or end detect*) /HMM menerima seluruh data percakapan dalam bentuk digital / biner, kemudian HMM akan melakukan proses konversi. Dalam proses konversi terdapat beberapa tahapan, diantaranya:
  - a) Data biner yang diterima oleh HMM akan dibandingkan dengan template data suara



Gambar 11. Perbandingan Data Biner dengan Template

- b) Tiap data biner yang telah dibandingkan dengan template suara, kemudian disatukan kembali dan dianalisis secara keseluruhan, kemudian akan dicocokkan dari segi tata bahasa dan apakah data yang diucapkan sesuai dengan kata yang tersedia pada template data pada dataset.

### SIMPULAN

Dari hasil penelitian yang telah dilakukan dapat ditarik kesimpulan bahwa.

1. *Library PocketSphinx* dapat digunakan untuk melakukan perubahan suara menjadi teks dengan akurasi kata sebesar 93,3%
2. Aplikasi dapat berjalan tanpa koneksi internet ketika melakukan pencarian kata.
3. Aplikasi speech ini dapat digunakan sebagai alat bantu untuk menterjemahkan bahasa Inggris, Indonesia dan Gorontalo.

4. Aplikasi ini menggunakan teknologi Speech To Text sebagai sarana pelatihan pengucapan yang tepat.

### DAFTAR PUSTAKA

- [1] Fahri Firdausillah, Ika Novita Dewi, Catur Supriyanto. Sphinx-4 Indonesian Isolated Digit Speech Recognition. *Journal of Theoretical and Applied Information Technology*. Volume 53.No.1. Halaman 40. 2013.
- [2] Safaat H., N. 2012. Pemrograman Aplikasi Mobile Smartphone dari Tablet PC Berbasis Android. Bandung: Informatika.
- [3] Atik Charisma. Sistem Verifikasi Penutur Menggunakan Metoda Mel Frequency Cepstral Coefficients-Vector Quantisation (Mfcc-Vq) Serta Sum Square Error (Sse) Dan Pengenalan Kata Menggunakan Metoda Logika Fuzzy. *Jurnal Teknik Elektro ITP*. Volume 2. No. 2. Halaman 24-25. 2013.
- [4] Danny Sulestio, Mayer Aristo, Linda Wijaya. Perancangan Aplikasi Pengenalan Suara untuk Menjalankan Beberapa Perintah Windows dengan Metode Back Propagation. *Jurnal Publikasi Ilmiah Binus*. Halaman 7. 2004.
- [5] Rabiner LR. 1989. *A Tutorial in Hidden Markov Models and Selected Applications in Speech Recognition*. *IEEE* 1989; 77: 257-287.