

KONVERTER SUARA DENGAN INPUT BAHASA INDONESIA KE VIDEO GERAKAN BAHASA ISYARAT DENGAN METODE *SPEECH RECOGNITION (HIDDEN MARKOV MODEL)* UNTUK PENDERITA TUNARUNGU

Resa Septiari¹, Hanny Haryanto²

²¹Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang 50131
E-mail :resa.septiari@gmail.com, 77project@gmail.com

ABSTRAK

Penelitian ini menjelaskan tentang implementasi speech recognition menjadi video bahasa isyarat untuk membantu penderita tunarungu dalam mengikuti kegiatan seminar dan ibadah. Untuk mengimplementasikan speech recognition, peneliti menggunakan bahasa pemrograman Visual Basic 6.0 dan untuk pengenalan suaranya menggunakan Microsoft Speech SDK. Untuk pengolahan suara menggunakan Hidden Markov Models (HMM) dengan Hidden Markov Toolkit (HTK) atau SPHINX. Pengolahan bahasa alami (Natural Language Processing) juga digunakan untuk mendapatkan hasil yang akurat dan agar dapat dimengerti oleh penderita tunarungu. Kata-kata yang diperoleh akan menjadi masukan ke pencarian video bahasa isyarat. Bahasa isyarat yang digunakan adalah American Sign Language (ASL) yang sama dengan Sistem Bahasa Isyarat Indonesia (SIBI).

Kata kunci : Speech Recognition, HMM, NLP, bahasa isyarat

1. PENDAHULUAN

1.1. Latar Belakang

Seperti yang kita ketahui, banyak informasi dari berbagai media yang dapat menambah pengetahuan kita. Kita mendapatkan informasi dari melihat dan mendengar, tapi masih banyak saudara kita yang belum bisa mendapatkan informasi secara maksimal karena adanya keterbatasan dari mereka. Ketika kita bisa mendengarkan seminar tentang informasi terbaru atau ceramah tentang keagamaan, saudara kita yang tunarungu mungkin akan kesulitan dalam menangkap apa yang dibicarakan oleh pembicara. Beberapa kegiatan seminar dan kegiatan keagamaan memang sudah memberikan solusi dengan meminta bantuan seorang interpreter bahasa isyarat (orang yang mengerti bahasa isyarat) untuk menerjemahkan bahasa biasa ke bahasa isyarat, namun hal ini masih kurang efektif karena terkadang sulit mencari orang yang mengerti bahasa isyarat dan menjadi interpreter.

Dengan berkembangnya teknologi seperti saat ini, telah dikembangkan beberapa software untuk memudahkan orang tunarungu dalam berkomunikasi dengan yang lain misalnya: Portable Sign Language Translator (PSLT) merupakan perangkat lunak yang dapat dibenamkan di perangkat portable untuk menerjemahkan bahasa isyarat ke dalam pesan tertulis yang dikembangkan oleh ilmuwan skotlandia. Kebanyakan dari perangkat lunak penunjang komunikasi tunarungu adalah bagaimana cara orang normal dapat mengerti komunikasi tunarungu (bahasa isyarat) atau memudahkan orang tunarungu berkomunikasi satu sama lain. Namun belum ada suatu perangkat lunak yang digunakan untuk mengkonversi suara pembicara ke bahasa isyarat dengan input bahasa Indonesia. Padahal orang tunarungu juga perlu mendapatkan informasi yang berasal media audio/ suara. Mereka biasanya mengandalkan interpreter jika di Indonesia biasanya yang menjadi interpreter adalah anggota keluarga mereka sendiri.

Sistem pengenalan wicara atau lebih dikenal dengan speech recognition dapat digunakan dalam mengatasi masalah ini namun permasalahannya adalah kebanyakan sistem pengenalan wicara menggunakan bahasa Inggris dan itupun terkadang ketepatannya masih kurang. Sistem pengenalan wicara dalam bahasa Indonesia masih terus dikembangkan sampai saat ini. Dari pengenalan wicara data yang didapat juga akan diolah menggunakan NLP (Natural Processing Language) suatu metode yang digunakan untuk mengolah data menjadi bahasa alami (bahasa yang dimengerti manusia, kasus disini: bahasa isyarat). NLP juga masih dalam perkembangan.

Berdasarkan uraian di atas maka peneliti mengambil topik dengan judul "Konverter Suara dengan Input Bahasa Indonesia ke Video Gerakan Bahasa Isyarat Menggunakan Metode Speech Recognition (Hidden Markov Model) untuk Penderita Tunarungu." yang diharapkan dapat membantu saudara kita yang memiliki keterbatasan pendengaran dalam mendapatkan informasi.

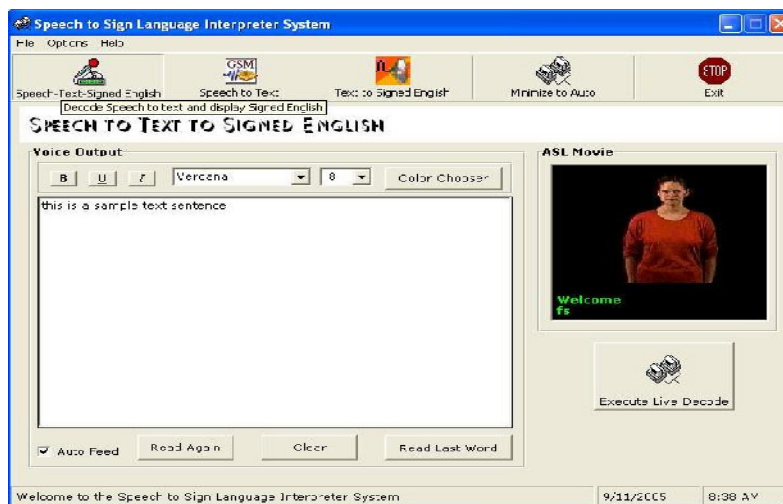
1.2. Tujuan

Tujuan pembuatan “Konverter Suara dengan Input Bahasa Indonesia ke Video Gerakan Bahasa Isyarat Menggunakan Metode Speech Recognition (Hidden Markov Model) untuk Penderita Tunarungu.” adalah:

1. Membantu saudara kita yang memiliki keterbatasan pendengaran dalam mendapatkan informasi.
2. Memudahkan para penyelenggara seminar atau acara lainnya dalam menyampaikan informasi kepada orang tunarungu dan tidak perlu mencari seorang interpreter.
3. Mengembangkan teknologi pengenalan wicara atau speech recognition dalam bahasa Indonesia dengan ketepatan yang tinggi.
4. Mengembangkan teknologi NLP (Natural Processing Language).

2. Tinjauan Pustaka

Konversi suara ke video bahasa isyarat untuk penderita tunarungu atau keterbatasan pendengaran adalah suatu perangkat lunak yang digunakan untuk membantu para penderita tunarungu dalam berkomunikasi dengan orang normal. Penelitian ini sudah pernah dilakukan oleh Khalid Khalil El-Darymli dkk dari International Islamic University Malaysia, ECE Dept., Faculty of Engineering dengan judul Speech to Sign Language Interpreter System (SSLIS). Dalam penelitian yang dilakukan oleh El-Darymli, bahasa suara yang digunakan (input) adalah bahasa Inggris dan video bahasa isyarat yang digunakan (output) adalah bahasa American Sign Language(ASL) dengan Speech Recognition(SR) engine Sphinx 3.5 dari Carnegie Mellon University[1].



Gambar 1. Interface dari Speech to Sign Language Interpreter System

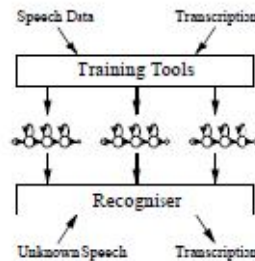
Bahasa isyarat adalah bahasa yang mengutamakan komunikasi manual, bahasa tubuh, dan gerak bibir untuk berkomunikasi. Kaum tunarungu adalah kelompok utama yang menggunakan bahasa ini, biasanya dengan mengombinasikan bentuk tangan, orientasi dan gerak tangan, lengan, dan tubuh, serta ekspresi wajah untuk mengungkapkan pikiran mereka. Bahasa Isyarat di Indonesia yang umum digunakan adalah *Sistem Isyarat Bahasa Indonesia (SIBI)* yang sama dengan American Sign Language(ASL) (Juniar Prima Rakhman, Translasi Bahasa Isyarat)[2].

Natural Language Processing (NLP) adalah pendekatan komputasi yang mampu menganalisis text yang berdasarkan sekumpulan teori dan teknologi. 'Teks Alami' dapat dari berbagai genre, modus bahasa, dll Teks-teks dapat lisan atau tertulis. Satu-satunya persyaratan yaitu teks tersebut adalah bahasa yang digunakan oleh manusia untuk berkomunikasi satu sama lain. (Liddy, E. D. In Encyclopedia of Library and Information Science, 2nd Ed. Marcel Decker, Inc)[3].

Speech Recognition adalah suatu sistem pengenalan suara yang dapat dikenali atau seolah-olah dapat dimengerti oleh komputer atau mesin dengan menggunakan algoritma atau tool tertentu. Salah satu tool yang digunakan untuk speech recognition adalah Hidden Markov Models. HMM adalah alat statistik yang digunakan untuk memodelkan urutan generatif yang dapat dicirikan oleh proses yang mendasari hasil dari observasi. HMM telah menemukan aplikasi di banyak bidang yang berkaitan dalam pemrosesan sinyal, dan khususnya speech processing, tetapi juga telah diterapkan dengan sukses pada NLP. Andrei Markov memberikan namanya kepada teori matematika Markov processes pada awal abad kedua

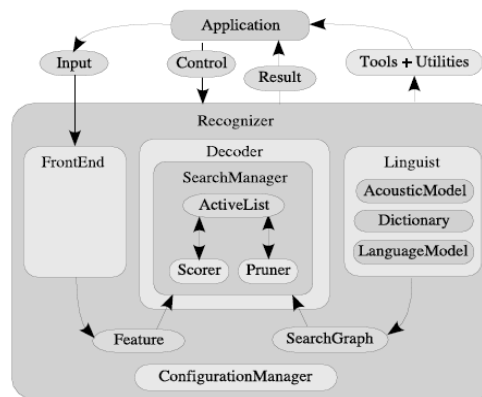
puluh, tapi setelah itu Baum dan rekan-rekannya yang mengembangkan teori HMMs pada tahun 1960 (Phil Blunsom, 2004)[4].

HTK adalah toolkit untuk membangun Model Hidden Markov (HMMs). HMM dapat digunakan untuk model pengenalan pola dan inti dari HTK sendiri juga untuk berbagai tujuan. Namun, tujuan utama HTK dirancang untuk membangun HMM-based speech processing tools di recognisers tertentu. Dengan demikian, banyak dukungan infrastruktur di HTK didedikasikan untuk ini. Seperti terlihat pada gambar 2, ada dua tahapan proses utama yang terlibat. Pertama, alat-alat pelatihan HTK digunakan untuk memperkirakan parameter dari serangkaian HMM menggunakan ucapan-ucapan pelatihan dan yang terkait transkripsi. Kedua, ucapan-ucapan yang tidak diketahui ditranskripsi menggunakan alat pengenalan HTK (Steve Young dkk,2009,The HTK Book Version 3.4)[5].



Gambar 2. Proses dari penggunaan HTK

Sedangkan SPHINX adalah speech recognition yang berbasiskan berbagai variasi dari HMM, yaitu Sphinx *discrete* HMM, *semicontinuous* HMM, dan *continuous* HMM. Saat ini versi terbaru dari SPHINX adalah SPHINX-4. Framework dari SPHINX-4 telah didesain dengan fleksibilitas dan modularitas yang tinggi.

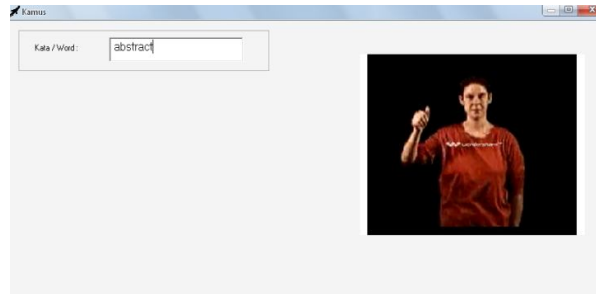


Gambar 3. Framework dari SPHINX-4

Gambar 3 menunjukkan keseluruhan arsitektur dari sistem. Setiap elemen yang berlabel adalah modul yang mudah diubah, mengizinkan peneliti untuk melakukan riset dengan menambahkan dan mengimplementasi modul yang berbeda. (Willie Walker dkk, 2004, Sphinx-4: A Flexible Open Source Framework for Speech Recognition)[6].

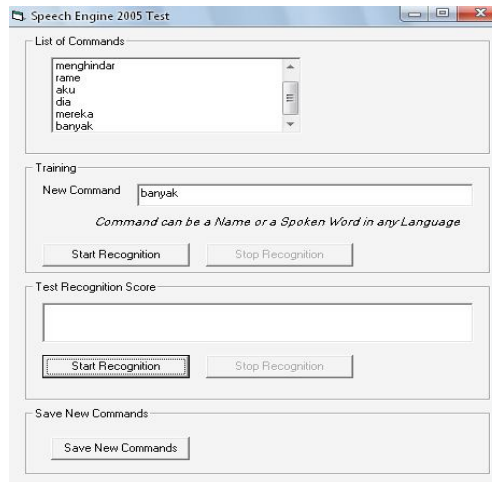
3. Metode Penelitian

Sebelum ini, peneliti telah melakukan dua kali penelitian dengan tema yang sama. Pertama dengan menggunakan bahasa Inggris sebagai input dan video ASL (*.swf) sebagai output dengan bahasa pemrograman Visual Basic 6.0 dan untuk speech recognitionnya menggunakan library dari Microsoft Speech Object Library dan Sistem Operasinya Windows XP sp 2. Ketepatan konverter menggunakan Microsoft Speech Object Library ini masih sangat rendah <50% sehingga hasil yang didapat belum memuaskan.



Gambar 4. Konverter Suara dalam Bahasa Inggris ke video ASL(*.swf) menggunakan bahasa VB 6.0

Penelitian kedua menggunakan bahasa Indonesia sebagai input dan video ASL (*.swf) sebagai output dengan bahasa pemrograman Visual Basic 6.0 dan untuk speech recognitionnya menggunakan Speech.sdk dari Research-Lab Small Foot-Print Speech Engine dan Sistem Operasinya Windows XP sp 2. Sebenarnya dengan tools dari Research-Lab ini, peneliti sudah dapat membuat suatu konverter bahasa Indonesia ke bahasa Isyarat dengan ketepatan yang cukup bagus sekitar 66% dan melalui proses training namun sayangnya tools ini bersifat komersial dan dibatasi penggunaannya.

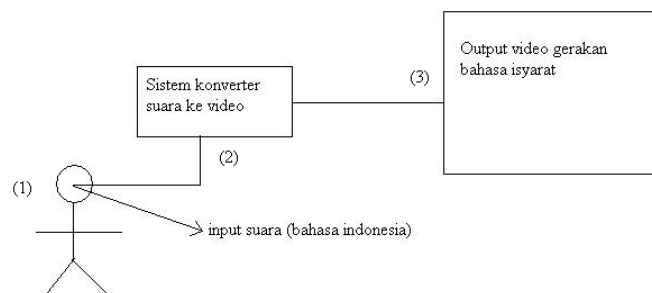


Gambar 5. Proses training dari Speech Engine Research-Lab

Maka dari itu, peneliti melakukan penelitian lagi dengan menggunakan metode dan bahasa yang berbeda. Metode yang peneliti lakukan untuk penelitian ini adalah:

1. Perencanaan sistem

Penjelasan mengenai alur proses kerja dari sistem “Konverter Suara dengan Input Bahasa Indonesia ke Video Gerakan Bahasa Isyarat Menggunakan Metode Speech Recognition (Hidden Markov Model) untuk Penderita Tunarungu” tampak pada gambar. Pembicara memberikan input suara dalam bahasa Indonesia (1). Sistem akan mengenali input suara dengan mencari ke database suara, jika ditemukan maka file video yang tersimpan akan segera dimainkan, jika ternyata suara tidak dikenali maka akan muncul suatu alert bahwa suara tidak dikenali(2). Sistem akan memberikan output berupa video gerakan bahasa isyarat (3).



Gambar 6. Alur proses kerja dari sistem “Konverter Suara dengan Input Bahasa Indonesia ke Video Gerakan Bahasa Isyarat Menggunakan Metode *Speech Recognition (Hidden Markov Model)* untuk Penderita Tunarungu”

Dalam mengenali suara sistem ini menggunakan SRM(*Speech Recognition Module*). SRM mengambil bagian pada interaksi user dengan microphone dimana suara yang dimasukkan kedalam microphone kemudian diekstraksi cirinya dan dibandingkan dengan *acoustic model* dimana *acoustic model* ini adalah kumpulan angka-angka parameter suara yang didapat HTK (*Hidden Markov Model Toolkit*) atau SPHINX, sedangkan Hidden Markov Toolkit dan SPHINX sendiri adalah trainer yang digunakan untuk mentrain SR engine dengan melakukan inputan berupa kata-kata dimana kata-kata ini belum pernah di inputkan sebelumnya kepada system dengan melakukan pembangunan HMM (*Hidden Markov Model*).

HMM terdiri atas *hidden state* (tidak dapat diamati/*hidden*) dan *feature vector* (dapat diamati/*observable*). Pembangunan model berarti pembangunan data probabilitas transisi antar-*hidden state* serta data probabilitas emisi (*emission*) yaitu pembangkitan *feature vector* oleh *hidden state*. Model HMM dapat dilihat pada Gambar 6. Konsep Hidden Markov Model sendiri adalah memodelkan simbol yang didapat atau diinputkan (disini berupa kata) ke sebuah mesin finite state, sehingga dapat diketahui simbol apa yang dapat mewakili sebuah parameter vektor dari input tersebut[5].

Dimisalkan O adalah sequence speech vector dari input suara.

$$O = o_1; o_2; \dots; O_t \quad (1)$$

Dimana O_t adalah speech vector yang diobservasi dalam waktu t .

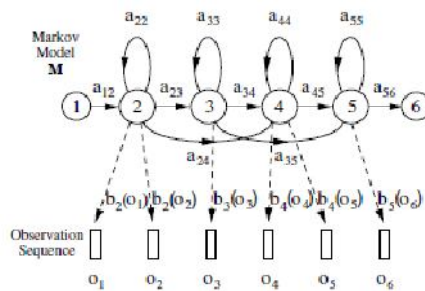
Pengenalan kata dapat ditentukan menggunakan persamaan:

$$\underset{i}{\operatorname{arg\,max}} \{P(w_i|O)\} \quad (2)$$

Dimana w_i adalah pengucapan yang ke- i , probabilitas ini tidak dapat dihitung secara langsung tetapi dihitung dengan aturan Bayes.

$$P(w_i|O) = \frac{P(w_i, O)}{P(O)} \quad (3)$$

Maka, prioritas kemungkinan $P(w_i)$ bergantung pada $P(O/w_i)$.



Gambar 7. Markov Generation Models

Speech recognition engine menggunakan Teknik pemotongan N-Grams. Dimana proses pemotongan ini dilakukan berdasarkan *Acoustic Model* yang dipakai. Dari hasil pemotongan kata yang diperoleh dari proses N-Gram ini kemudian dicari kemungkinan kemiripannya dengan kata yang terdapat dalam *dictionary*, Setelah didapatkan hasil yang paling mirip, hasil inilah yang menjadi hasil dari SRM. Setelah diperoleh hasil output kemudian hasil ini di inputkan dalam pencarian video ke dalam Visual Basic 6.0 untuk dilakukan proses selanjutnya.

2. Pengumpulan Bahan dan Source

Pengumpulan bahan berupa rekaman menggunakan wavesurfer dan Hidden Markov Model Toolkit (HTK) atau SPHINX. HTK adalah tool atau perangkat lunak yang mudah digunakan untuk membangun dan memanipulasi HMM. HTK pada dasarnya digunakan untuk penelitian mengenai pengenalan suara meskipun juga digunakan untuk sejumlah aplikasi yang lain termasuk penelitian dalam suara sintesis, pengenalan karakter, dan pengurutan DNA. HTK terdiri dari kumpulan beberapa modul library dan tool dalam bahasa C. Perangkat-perangkat tersebut memberikan fasilitas yang handal untuk *speech analysis*, HMM *training*, testing dan results analysis. Sedangkan SPHINX hampir sama dengan HTK tool untuk mengenali suara berbasis HMM. Desain dari SPHINX-4 adalah perkembangan dari sistem-sistem

yang terdahulu dan memiliki komabilitas yang jauh lebih baik. Pengumpulan video bahasa isyarat untuk sementara didapat dari website resmi *American Sign Language* (ASL).

3. Instalasi source

Secara keseluruhan hampir semua proses instalasi tools dapat bekerja secara tepat dan baik, sesuai dengan keinginan peneliti namun ada beberapa tools yang tidak berjalan dengan baik pada platform Windows seperti HTK yang beberapa fungsi belum berjalan secara normal di Windows.

4. Hasil dan Pembahasan Sementara

1. Menggunakan Bahasa Inggris sebagai Input dan Video ASL (*.swf) sebagai Output.

Pengujian program ini melibatkan lima orang yaitu dua wanita(W) dan tiga pria(P). Pengujian ini juga menggunakan tiga sample input kata yaitu: *many, you, about*.

Tabel 1: Hasil Pengujian Kata *Many*

Pengujian	Teks yang dihasilkan	Video
W1	many	v
W2	winning	x
P1	many	v
P2	been	x
P3	winning	x

Tabel 2: Hasil Pengujian Kata *You*

Pengujian	Teks yang dihasilkan	Video
W1	you	v
W2	year	x
P1	you	v
P2	use	x
P3	you	v

Tabel 3: Hasil Pengujian Kata *About*

Pengujian	Teks yang dihasilkan	Video
W1	bill	x
W2	about	v
P1	about	v
P2	the	x
P3	Bill	x

Dari data di atas diperoleh hasil akurasi *speech to text* sebesar 46.7%. Sedangkan untuk konversi teks ke video sudah tidak mengalami masalah.

2. Menggunakan Bahasa Indonesia sebagai Input dan Video ASL (*.swf) sebagai Output.

Pengujian program ini melibatkan lima orang yaitu dua wanita(W) dan tiga pria(P). Pengujian ini juga menggunakan tiga sample input kata yaitu: *aku, dia, mereka*.

Tabel 1: Hasil Pengujian Kata *Aku*

Pengujian	Teks yang dihasilkan	Video
W1	aku	v
W2	aku	v
P1	aku	v
P2	hari	x
P3	aku	v

Tabel 2: Hasil Pengujian Kata *Dia*

Pengujian	Teks yang dihasilkan	Video
W1	dia	v
W2	iya	x
P1	dia	v
P2	dia	v
P3	bisa	x

Tabel 3: Hasil Pengujian Kata *Mereka*

Pengujian	Teks yang dihasilkan	Video
W1	menerka	x
W2	mereka	v
P1	merk	x
P2	mereka	v
P3	mereka	v

Dari data di atas diperoleh hasil akurasi *speech to text* sebesar 66.67%. Sedangkan untuk konversi teks ke video sudah tidak mengalami masalah.

5. Penutup

Peneliti sedang melakukan eksperimen untuk membuat suatu SR modul bahasa Indonesia berbasis HMM yang dapat digunakan dalam bahasa pemrograman Visual Basic. Sebelumnya peneliti telah melakukan dua eksperimen dengan hasil yang kurang memuaskan. Kendala utama dalam melakukan penelitian ini adalah pengenalan suara ke sistem menjadi teks sedangkan untuk konversi dari teks ke video bahasa isyarat tidak mengalami kendala yang berarti dari sisi programingnya. Namun minimnya sumber video yang berhasil didapat oleh peneliti membuat peneliti menggunakan video yang diambil dari web resmi ASL.

Perbaikan yang dapat dilakukan dalam program ini adalah menambah frekuensi proses *speech training* sehingga akurasi yang dihasilkan pada pengenalan suara dapat lebih tinggi. Pada situasi yang bising, pengenalan suara juga tidak dapat bekerja secara maksimal. Hal ini tentu saja akan berpengaruh pada output video yang dihasilkan.

DAFTAR PUSTAKA

- [1] Khalid Khalil El-Darymli, Othman O. Khalifa and Hassan Enemosah, "Speech to Sign Language Interpreter System (SSLIS)," *International Islamic University Malaysia, ECE Dept., Faculty of Engineering*
- [2] Juniar Prima Rakhman, *Translasi Bahasa Isyarat*.
- [3] Liddy, E. D, "Encyclopedia of Library and Information Science, 2nd Ed," Marcel Decker, Inc.
- [4] Phil Blunsom, "Hidden Markov Models," 2004.
- [5] Steve Young dkk, "The HTK Book Version 3.4," 2009.
- [6] Willie Walker dkk, "Sphinx-4: A Flexible Open Source Framework for Speech Recognition," 2004