

APLIKASI PENGENALAN SUARA MENGGUNAKAN MICROSOFT SAPI SEBAGAI PENGENDALI PERALATAN ELEKTRONIK

Sukarso¹, Abdusy Syarif²

Teknik Informatika, Universitas Mercu Buana, Jakarta
e-mail: ¹sukarso@gmail.com, ²abdusyarif@mercubuana.ac.id

ABSTRAKSI

Perintah dengan menggunakan suara memungkinkan pengguna dapat mengontrol sebuah peralatan elektronik dengan mengucapkan perintah melalui peralatan masukan audio daripada harus menggunakan saklar, hal ini memberikan kemudahan bagi pengguna untuk mengontrol peralatan elektronik tanpa menggunakan tangan.

Perintah dengan suara menggunakan peralatan masukan audio seperti mikropon, sebuah mesin pengenalan ucapan dan menu perintah. Ketika pengguna mengucapkan perintah melalui mikropon, mesin pengenalan ucapan akan menterjemahkan dari ucapan ke teks. Jika mesin pengenalan ucapan menemukan kecocokan perintah suara dengan menu perintah suara maka akan memerintahkan aplikasi yang cocok untuk dijalankan. Perintah dengan menggunakan suara merupakan *speaker independent* (tidak terpengaruh siapa yang mengucapkan), jadi tidak diperlukan pelatihan suara dari pengguna.

Untuk dapat mengontrol peralatan elektronik, maka aplikasi ini memerlukan mikroprosesor AT89C51 sebagai pengendali untuk mengaktifkan saklar relay menggunakan komunikasi port serial.

Masalah yang dihadapi dalam menggunakan pengenalan ucapan adalah kebisingan diperlukan mikropon yang berkualitas tinggi. Juga perlu diingat untuk berbicara sejelas mungkin dan dalam ucapan tempo yang sedang (tidak terlalu cepat/lambat). Bercampurnya suara yang tidak diinginkan serta ucapan yang terlalu cepat dapat menyebabkan ketidakakuratan dalam pengenalan ucapan.

Kata kunci: Perintah Suara, Pengenalan Ucapan, SAPI

1. PENDAHULUAN

Dewasa ini teknologi pengenalan suara merupakan teknologi yang menjanjikan di masa depan, contoh yang signifikan dari teknologi ini adalah aplikasi *voice command* yang diharapkan bisa mengatasi banyak aspek dalam kehidupan sehari-hari. Sebagai contoh bagi saudara-saudara kita yang mengalami keterbatasan fisik dapat memanfaatkan aplikasi berbasis teknologi pengenalan suara untuk menyalakan peralatan elektronik dengan bantuan komputer yang telah dilengkapi dengan aplikasi *voice command*.

Tujuan dari penelitian ini adalah sebagai berikut:

- Membuat sebuah perangkat lunak menggunakan teknologi pengenalan suara yang dapat mengontrol peralatan elektronik.
- Memberikan salah satu pilihan dalam mengimplementasikan pengontrolan peralatan elektronik di rumah dengan memberikan perintah berupa ucapan/suara selain dengan cara yang sudah ada dengan menggunakan saklar yang saat ini kita kenal.

Fungsi utama dari perangkat lunak yang akan dikembangkan adalah perintah menggunakan suara (*voice command*). Secara garis besar, keseluruhan proses yang dilakukan oleh perangkat lunak ini ada dua buah proses yaitu:

- Proses pengenalan suara yang diucapkan

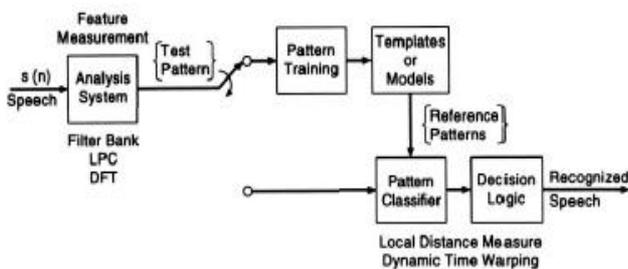
pengguna (user) oleh komputer.

- Ucapan yang telah dikenali digunakan untuk mengontrol peralatan elektronik dengan menggunakan pengendali mikroprosesor AT89C51. Mikroprosesor AT89C51 merupakan mikroprosesor yang dapat mengolah data per bit ataupun 8 bit yang dikirim oleh komputer melalui jalur serial port.

Pembahasan perangkat lunak yang dibangun hanya pada aplikasi pengenalan ucapan saja, tidak termasuk pada rancangan modul pengendali.

2. SISTEM PENGENAL UCAPAN (SPEECH RECOGNITION SYSTEM)

Sistem Pengenalan Ucapan (*Speech Recognition System*) adalah sistem yang berfungsi untuk mengubah bahasa lisan menjadi bahasa tulisan [1]. Masukan sistem adalah ucapan manusia, selanjutnya sistem akan mengidentifikasi kata atau kalimat yang diucapkan dan menghasilkan teks yang sesuai dengan apa yang diucapkan. Gambar 1 diperlihatkan konfigurasi tipikal suatu sistem pengenalan ucapan [2]. Sinyal ucapan pertama kali akan dilewatkan pada bagian penganalisis ucapan untuk mendapatkan besaran-besaran atau ciri-ciri yang mudah diolah pada tahap berikutnya. Untuk setiap ucapan yang berbeda akan dihasilkan pola ciri yang berbeda.



Gambar 1. Diagram Blok Sistem Pengenalan Ucapan [2]

Penganalisis sintaks akan melakukan transformasi sinyal ucapan dari domain waktu ke domain frekuensi. Pada domain frekuensi, untuk kurun waktu yang singkat, setiap sinyal dapat terlihat memiliki ciri-ciri yang unik. Namun demikian, pengucapan suatu unit bunyi ucapan (fonem) seringkali bervariasi antar orang yang berbeda, juga terpengaruh oleh fonem-fonem disekitarnya, kondisi emosi, noise, dan faktor-faktor lainnya. Sistem *Speech Recognition* akan melakukan pengenalan untuk setiap unit bunyi pembentuk ucapan (fonem), selanjutnya mencoba mencari kemungkinan kombinasi hasil ucapan yang paling dapat diterima. Sistem yang lebih sederhana adalah sistem yang hanya dapat mengenal sejumlah kata yang jumlahnya terbatas. Sistem ini biasanya lebih akurat dan lebih mudah dilatih, tetapi tidak dapat mengenal kata yang berada di luar kosa kata yang pernah diajarkan.

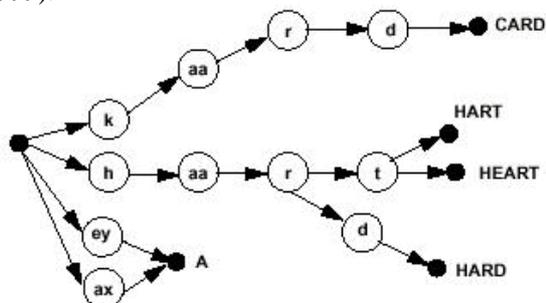
Ada 2 tipe *Speech Recognition* [3], dilihat dari ketergantungan pembicara yaitu:

- Independent Speech Recognition*, yaitu sistem pengenalan ucapan tanpa terpengaruh dengan siapa yang berbicara, tetapi mempunyai keterbatasan dalam jumlah kosakata. Model ini akan mencocokkan setiap ucapan dengan kata yang dikenali dan memilih yang "sepertinya" cocok. Untuk mendapatkan kecocokan kata yang diucapkan maka digunakan model statistik yang dikenal dengan nama *Hidden Markov Model* (HMM)
- Dependent Speech Recognition*, yaitu sistem pengenalan ucapan yang memerlukan pelatihan khusus dari pembicara, dimana hasil pelatihan dari masing-masing pembicara akan disimpan dalam sebuah profil. Profil inilah yang nantinya digunakan untuk berinteraksi dengan sistem pengenalan ucapan dan sistem akan bergantung siapa yang berbicara. Sistem ini biasanya lebih mudah untuk dikembangkan, dimana contoh suara sudah dibuat sebelumnya dan disimpan dalam *database* (basis data) dan jumlah kosakatanya lebih besar dibandingkan dengan *independent speech recognition*. Proses pengenalan ucapan dengan cara membandingkan ucapan pembicara dengan

contoh suara yang ada.

2.1 Model Statistik untuk Pengenal Ucapan

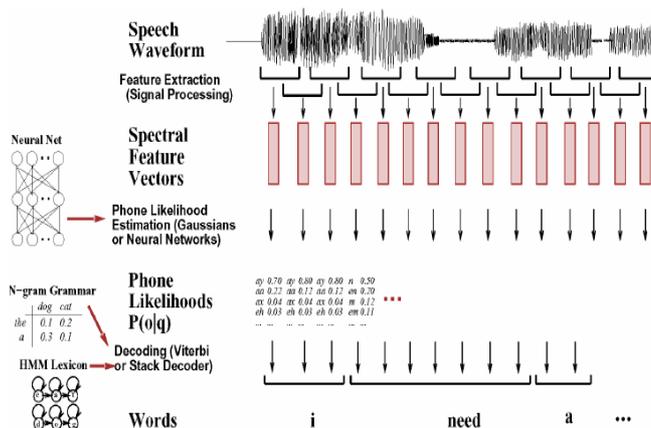
Saat ini sistem pengenalan ucapan telah menggunakan model statistik untuk menentukan dengan tepat kata yang dimaksud. Menurut John Garofolo, *Speech Group Manager* pada *Information Technology Laboratory* di *National Institute of Standards and Technology*, model statistik yang sangat mendominasi adalah *Hidden Markov Model* dan *Neural Networks*. Tetapi yang paling sering digunakan adalah *Hidden Markov Model*, dimana setiap fonem dihubungkan seperti rantai, dan rangkaian rantai yang komplit merupakan kata. Gambar 3. memperlihatkan Model Statistik berdasarkan *Hidden Markov Model* (*IEEE Signal Processing Magazine*, vol. 16, no. 5, September 1999).



Gambar 2. Model Statistik Fonem Berdasarkan Hidden Markov Model [4].

2.2 Tahapan Pengenalan Ucapan

Secara teknis, Sistem Pengenalan Ucapan dapat dijabarkan menggunakan contoh seperti pada Gambar 3. (MIT Lecture Notes: *Introduction to Automatic Speech Recognition*, Lecture #1, Session 2003).



Gambar 3. Contoh Penjabaran Teknis Sistem Pengenalan Ucapan [5].

Sumber sinyal yang berasal dari ucapan seseorang, dinyatakan sebagai **Speech Waveform**. Sumber sinyal ini kemudian dicari ciri pentingnya yang dapat membedakan antara satu sinyal dengan

sinyal lain. Pengungkapan ciri dilakukan dengan pengolahan sinyal [5].

Hasil dari pengolahan dan analisis sinyal adalah deretan vektor ciri ucapan, dinyatakan sebagai **spectral feature vector**. Vektor-vektor ini dilatih menggunakan Jaringan Syaraf Tiruan. Pelatihan dimaksudkan agar pemrosesan vektor dapat dilakukan dengan cepat. Hasil dari pelatihan berupa deretan fonem-fonem yang membentuk data ucapan. Pemodelan fonem ini menggunakan model akustik *Hidden Markov Model* (HMM).

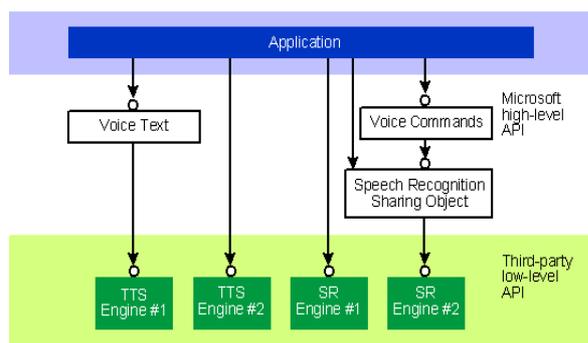
Proses pencarian kesamaan data fonem dilakukan dengan menggunakan **Grammar N-Gram** dan dikodekan kembali menggunakan algoritma **Viterbi**. Proses ini melihat nilai total penjumlahan terbaik. Dari penjumlahan ini diperoleh deretan kata yang terpilih dan diasumsikan yang paling benar, tampak hasil **"I need a"** pada Gambar 4.

2.3 Microsoft SAPI

Speech Application Programming Interface (SAPI) adalah sebuah API yang dikembangkan oleh Microsoft yang digunakan sebagai pengenalan suara didalam lingkungan pemrograman aplikasi Windows. Sampai saat ini SAPI dikemas baik berupa SDK (*System Development Kit*) maupun disertakan dalam sistem operasi Windows itu sendiri. Aplikasi yang telah menggunakan SAPI antara lain *Microsoft Office*, dan *Windows Vista*.

Secara arsitektur pemrograman SAPI dapat dilihat sebagai sebuah *middleware*[6] yang terletak antara aplikasi dan *speech engine*.

Di dalam SAPI versi 1 sampai dengan 4, aplikasi dapat berkomunikasi langsung dengan *speech engine* seperti tampak pada Gambar 4.



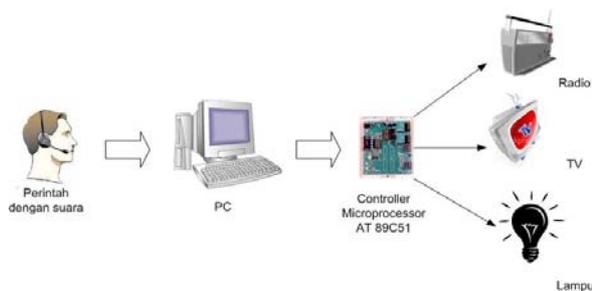
Gambar 4. Arsitektur Microsoft SAPI [6]

Komponen utama di dalam SAPI 4 adalah sebagai berikut:

- Voice Command*, sebuah obyek level tinggi untuk perintah dan kontrol menggunakan pengenalan suara.
- Voice Dictation*, sebuah obyek level tinggi untuk *continous dictation speech recognition*.
- Voice Talk*, sebuah obyek level tinggi untuk *speech synthesis*.

- Voice Telephony*, sebuah obyek untuk menulis aplikasi telepon berbasis pengenalan suara.
- Direct Speech Recognition*, sebuah obyek sebagai mesin untuk mengontrol pengenalan suara (*direct control of recognition engine*)
- Direct Text to Speech*, sebuah obyek sebagai mesin yang mengontrol *synthesis*.
- Audio Object*, untuk membaca dari audio device atau sebuah file audio.

3. ANALISIS DAN PERANCANGAN



Gambar 5. Arsitektur Aplikasi Pengenalan Suara Menggunakan Microsoft SAPI Sebagai Pengendali Peralatan Elektronik.

Dari Gambar 5 dapat dijelaskan sebagai berikut:

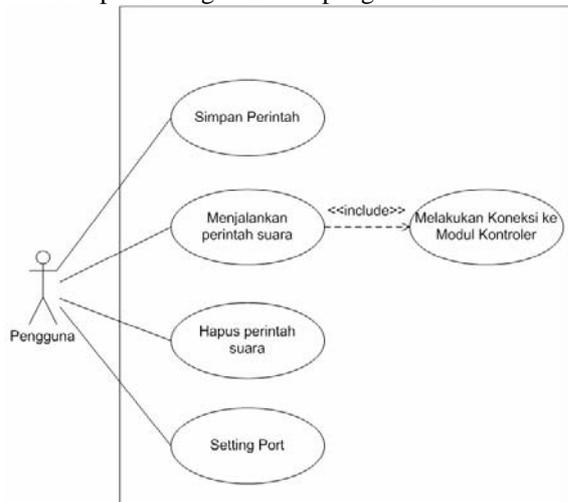
- Perintah berupa suara manusia akan diterima oleh mikropon.
- Dari mikropon sinyal suara diubah menjadi sinyal analog.
- Sinyal analog diterima oleh komputer melalui kartu suara (*sound card*) dan dirubah menjadi sinyal digital.
- Dengan menggunakan *Microsoft Speech Application Protocol Interface*, suara akan dibagi-bagi menjadi fonem. Dari fonem akan disusun menjadi sebuah kata yang cocok dengan data didalam berkas teks.
- Kata yang cocok akan digunakan untuk memerintahkan modul pengendali AT89C51 melalui Port Serial menggerakkan saklar relay didalam modul pengendali.
- Saklar relay akan menghidupkan dan mematikan peralatan elektronik yang terhubung ke listrik.

3.1 Diagram Use Case

Gambar 6 menggambarkan use case aplikasi yang dibangun. Use case yang ada di dalam Aplikasi Pengenalan Suara Menggunakan Microsoft SAPI Sebagai Pengendali Peralatan Elektronik adalah sebagai berikut:

- Simpan Perintah, digunakan untuk menyimpan perintah baru.
- Hapus Perintah, digunakan untuk menghapus perintah
- Setting Port, digunakan untuk melakukan perubahan konfigurasi serial port.

- d. Melakukan Koneksi ke Modul Pengendali, digunakan untuk melakukan koneksi antara komputer dengan modul pengendali.



Gambar 6. Diagram Use Case Sistem Pengenalan Suara Menggunakan Microsoft SAPI Sebagai Pengendali Peralatan Elektronik

3.2 Diagram Sekensial dan State

Diagram sekuensial untuk proses menjalankan perintah suara diperlihatkan pada Gambar 7 dimana penjelasan dari diagram sekuensial tersebut adalah sebagai berikut:

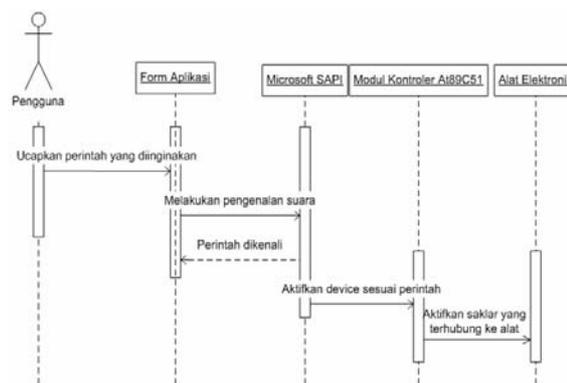
- Pengguna mengucapkan perintah.
- Microsoft SAPI akan mengenali suara yang diucapkan pengguna.
- Suara yang dikenali akan mengaktifkan device pengendali.
- Device pengendali akan menghidupkan peralatan elektronik.

Penjelasan Diagram Aktifitas aplikasi Sistem Pengenalan Suara sebagai Kontrol Peralatan Elektronik secara keseluruhan adalah sebagai berikut:

- Pengguna menjalankan aplikasi dan menunggu sampai form Aplikasi ditampilkan.
- Setelah form Aplikasi tampil, maka pengguna dapat memilih menu yang diinginkan.
- Jika memilih menyimpan perintah, maka pengguna harus memilih device yang akan diberi perintah, kemudian tekan tombol Save, maka perintah akan disimpan didalam berkas

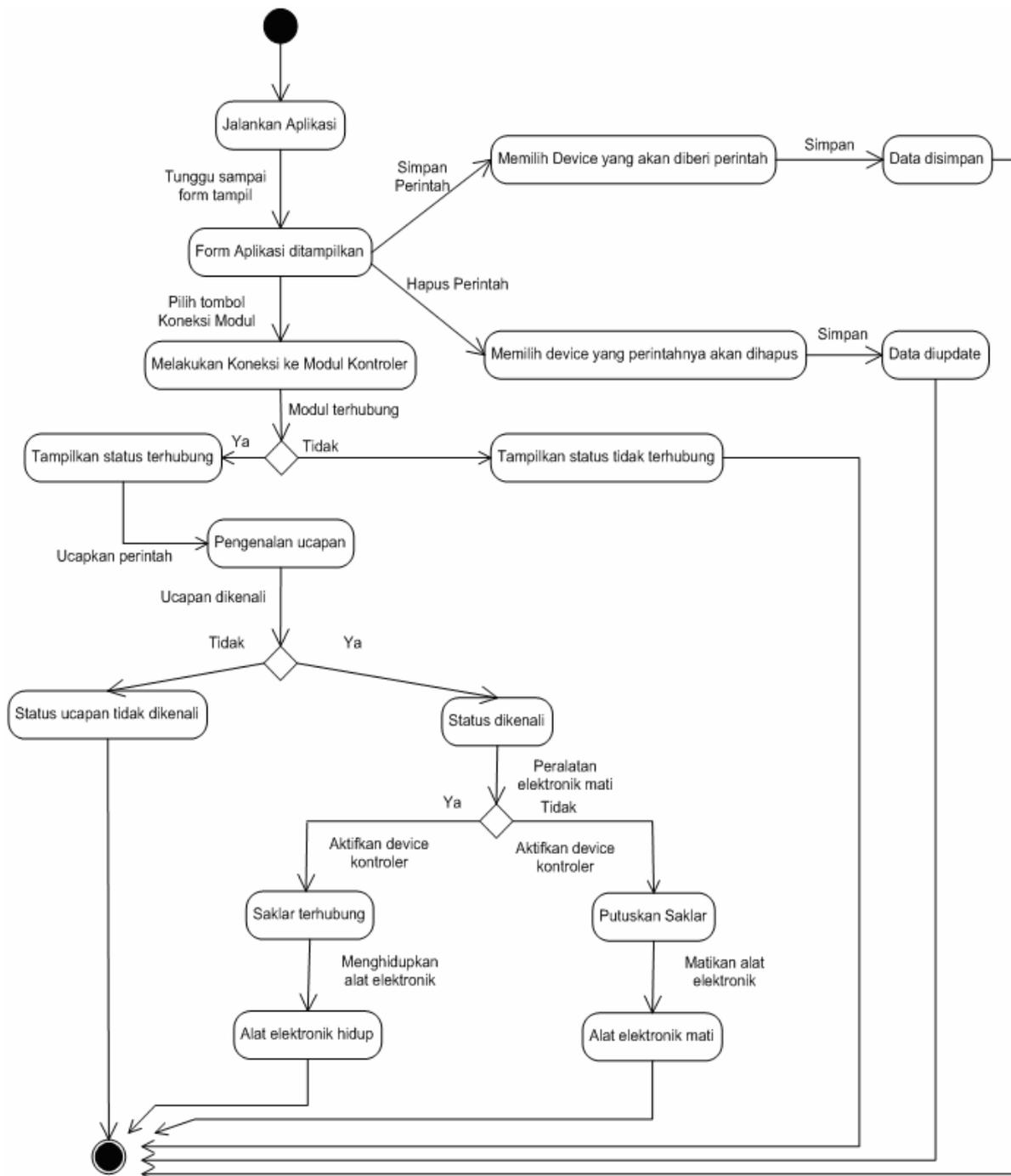
teks.

- Jika memilih menghapus perintah, maka pengguna harus memilih device dimana perintah akan dihapus, setelah dihapus kemudian tekan tombol Save untuk melakukan pemutakhiran data di dalam berkas teks.
- Jika memilih menu Port Setting maka pengguna akan melakukan perubahan terhadap setingan port, misalnya merubah dari Port 1 ke Port 2.



Gambar 7. Diagram Sekensial Menjalankan Perintah Suara

- Jika pengguna akan memberikan perintah berupa ucapan, maka pengguna harus melakukan koneksi terlebih dahulu ke modul pengendali.
- Setelah status modul terkoneksi, maka pengguna dapat memberikan perintah ucapan.
- Jika ucapan dikenali dan kondisi peralatan elektronik yang terhubung dengan device dalam kondisi mati, maka saklar otomatis yang terhubung ke modul akan terhubung dan peralatan elektronik akan hidup.
- Jika ucapan dikenali dan kondisi peralatan elektronik yang terhubung dengan device dalam kondisi hidup,
- maka saklar otomatis yang terhubung ke modul akan terputus dan peralatan elektronik akan mati.
- Jika ucapan tidak dikenali, maka status pesan tidak dikenali.



Gambar 8. Diagram State Aplikasi Sistem Pengenalan Suara Menggunakan Microsoft SAPI sebagai Pengendali Peralatan Elektronik

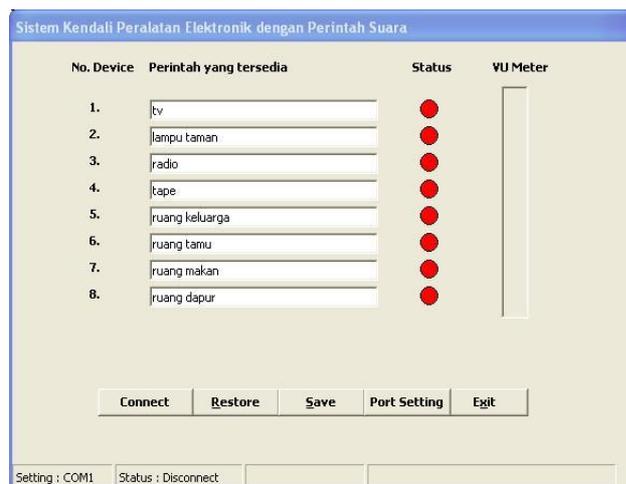
4. IMPLEMENTASI DAN PENGUJIAN

Lingkungan implementasi yang digunakan untuk mengembangkan perangkat lunak ini adalah sebuah komputer dengan spesifikasi:

- a. Perangkat keras komputer:
 - ✓ Processor Intel Pentium 4 M 2.4 GHz.
 - ✓ Memori 512 MB.
- b. Perangkat lunak komputer:
 - ✓ Sistem Operasi Windows XP Home Edition
 - ✓ Microsoft Visual Studio 6.0

- c. Perangkat Keras Modul Pengendali
 - ✓ Microprocessor buatan ATMEL AT89C51

Antarmuka utama merupakan form utama dari Aplikasi Pengenalan Suara Menggunakan Microsoft SAPI Sebagai Pengendali Peralatan Elektronik dimana semua perintah dan tombol-tombol yang digunakan untuk menjalankan aplikasi ini ada pada form ini seperti pada Gambar 9.



Gambar 9. Antarmuka Utama

Pengujian fungsional yang dilakukan pada perangkat lunak yang diimplementasikan pada Aplikasi Pengenalan Suara Menggunakan Microsoft SAPI Sebagai Pengendali Peralatan Elektronik meliputi beberapa skenario, yaitu: melakukan koneksi ke modul pengendali, menjalankan perintah suara, dan pengujian non fungsional untuk mengetahui pengaruh pengenalan suara untuk cepat/lambat dalam pengucapan.

Analisa pengujiannya adalah sebagai berikut:

- Untuk pengujian koneksi ke modul pengendali hasil dari pengujian telah sesuai dengan harapan.
- Pengujian untuk menjalankan perintah suara yang dilakukan oleh pengguna hasil dari pengujian menunjukkan lampu taman menyala dan status bar memberikan informasi status perintah yang dikenali.
- Pengujian terhadap pengaruh cepat/lambat untuk mengetahui bagaimana sistem dapat mengenali ucapan. Ketika pengguna mengucapkan perintah "Ruang Tamu" yang diucapkan dalam durasi 0,65 detik, sistem mengenali ucapan perintah tersebut. Begitu pula ketika pengguna memberikan perintah yang sama dengan durasi yang lebih lambat yaitu 1,95 detik. Dibawah 0,65 detik sistem mengalami kesulitan pada saat pengenalan ucapan, karena dengan durasi yang cepat fonem akan sulit di prediksi oleh mesin pengenalan ucapan. Untuk perintah ucapan yang diucapkan lebih lambat lagi dari 1,95 detik juga mengalami kesulitan dalam pengenalannya.

5. KESIMPULAN DAN SARAN

Beberapa kesimpulan yang dapat diambil dalam penelitian ini adalah sebagai berikut:

- Aplikasi Pengenalan Suara Menggunakan Microsoft SAPI Sebagai Pengendali Peralatan

Elektronik dapat dijadikan sebagai salah satu solusi dalam mengendalikan peralatan elektronik daripada menggunakan saklar secara langsung.

- Ada beberapa faktor yang mempengaruhi tingkat keberhasilan dalam menjalankan aplikasi Pengenalan Suara Menggunakan Microsoft SAPI sebagai Pengendali Peralatan Elektronik, yaitu:

- ✓ Kebisingan lingkungan. Lingkungan yang bising (noise) akan mempengaruhi keakuratan pengenalan ucapan.
- ✓ Tempo pengucapan perintah suara. Agar perintah yang diucapkan dapat mudah dikenali, maka pengguna harus mengucapkan perintah dengan jelas dan dengan tempo yang tidak terlalu cepat ataupun lambat.

Saran-saran yang dapat disampaikan adalah sebagai berikut:

- Perintah suara yang digunakan dalam Aplikasi Pengenalan Suara Menggunakan Microsoft SAPI sebagai Pengendali Peralatan Elektronik akan lebih akurat jika perintah menggunakan bahasa Inggris.
- Agar aplikasi ini dapat digunakan dalam jangkauan yang lebih jauh, perlu digunakan mikropon nirkabel (wireless), sehingga pengguna dapat melakukan kontrol peralatan elektronik dari jarak jauh.
- Aplikasi ini akan lebih akurat dalam pengenalannya jika digunakan pada lingkungan yang tenang (*quite environment*).

PUSTAKA

- Arman, Arry Akhmad. *Proses Pembentukan dan Karakteristik Sinyal Ucapan*. <http://indotts.melsa.net.id/Karakteristik%20Sinyal%20Ucapan.pdf>.
- Rabiner, Lawrence. 1993. *Fundamental of Speech Recognition*. Prentice Hall.
- Mains, Ron. 2001. *White Paper On Speech Recognition In The SESA Call Center*. University of Maryland.
- Deshmukh, Neeraj. Hierarchical Search for Large Vocabulary Conversational Speech. *IEEE Signal Processing Magazine*, Vol. 16, No. 5, September 1999.
- Banowosari, Lintang Yuniar, *Proposal "Adaptasi Sistem Pengenalan Ucapan Bahasa Inggris Ke Dalam Sistem Pengenalan Ucapan Bahasa Indonesia Baku Menggunakan Pendekatan Bootstrapping Termodifikasi"*. <http://staffsite.gunadarma.ac.id/lintang/>
- Rozak, Mike. *Talk to Your Computer and Have It Answer Back with the Microsoft Speech API*. <http://www.microsoft.com/msj/archive/S233.aspx>.