

# DEPENDENT DAN INDEPENDENT SPEAKER UNTUK KONTROL PERALATAN RUMAH TANGGA MELALUI PENGENALAN WICARA

Sri Utami

Jurusan Teknik Konversi Energi Politeknik Negeri Bandung

Email : [utamiwu@yahoo.com](mailto:utamiwu@yahoo.com)

## Abstrak

Konsumsi energi dapat dikurangi salah satunya dengan pengontrolan dan monitoring peralatan rumah tangga setiap waktu. Pengontrolan akan lebih efektif bila bisa dilakukan setiap saat dan melalui pengenalan wicara. Pengenalan wicara diimplementasikan untuk mengontrol dan memonitoring peralatan rumah tangga jarak jauh. Proses kontrol dilakukan dari PC client ke PC server. Sistem ini mempunyai input berupa dependent dan independent speaker. Independent speaker digunakan untuk mengetahui sistem bisa bekerja untuk speaker yang bukan merupakan speaker standar. Proses yang digunakan dalam penelitian ini adalah pengambilan sinyal, sampling, frame, windowing, LPC (Linear Prediction Code). Proses-proses tersebut dinamakan sebagai proses ekstraksi. Setelah itu dilakukan proses Dynamic Time Warping (DTW) untuk memadankan sinyal masukan dengan sinyal standar. Proses ekstraksi dan Dynamic Time Warping dilakukan pada PC client, hasilnya dikirim ke server untuk mengontrol alat yang diinginkan. Setelah alat terkontrol, perangkat keras akan memberikan feedback ke PC server kemudian mengirimkannya ke PC client. Di PC client dapat diketahui kondisi peralatan yang dikontrol. Alat yang dikontrol adalah empat buah lampu. Sistem yang dibangun mempunyai prosentase keberhasilan rata-rata 100% dependent speaker (pengucap yang suaranya dipakai sebagai standar) dan prosentase keberhasilan rata-rata 71,88% untuk independent speaker (pengucap yang suaranya tidak dipakai sebagai standar).

*Kata Kunci: dependent dan independent speaker, DTW*

## Abstract

Energy consumption can be reduced by controlling and monitoring household appliances at all times. Speech recognition was implemented to control and monitor home appliances remotely. The system based on dependent and independent speaker models. Independent speaker was used to determine the system works for speakers that are not standard. The process were taking signals, sampling, frames, windowing, LPC (Linear Prediction Code) and called the extraction process. Then Dynamic Time Warping (DTW) match the input signal with a standard signal. Extraction and Dynamic Time Warping process were on the client PC, the result was sent to the server to control the desired device. The hardware will provide feedback to the server PC and then sends it to the client PC. The condition of the equipment being controlled can be seen on the client PC. The controlled device are four lamps. The system has an average success percentage 100% for dependent speaker and 71.88% for independent speaker.

*Key Words: dependent dan independent speaker, DTW*

## PENDAHULUAN

Penghematan energi dapat dicapai dengan mengurangi konsumsi dan kegiatan yang menggunakan energi. Manfaat yang diperoleh sama tapi dengan tingkat efisiensi yang jauh lebih tinggi, bisa dilakukan salah satunya dengan mengontrol peralatan rumah tangga. Peralatan rumah tangga yang merupakan kebutuhan sehari-hari bila digunakan secara bijaksana akan menyebabkan berkurangnya biaya.

Penggunaan energi dikendalikan berdasarkan perintah yang dilakukan secara manual, *wireless*, *timer*, pengendalian jarak jauh, dan pengendalian jarak jauh dengan pengenalan wicara. *Intelligent personal assistant* dan *knowledge navigator* yang merupakan salah satu aplikasi yang ada pada iPhone 4S telah memanfaatkan efek suara untuk pengendalian. Dewantara<sup>[2]</sup> membuat aplikasi pengenalan wicara untuk perintah nirkabel robot mikro *mouse*, Anggraeni dan Astutik<sup>[1]</sup> menggunakan pengenalan wicara untuk mengakses suatu jaringan sebagai *password*.

Di samping itu, dalam beberapa tahun terakhir ini perkembangan dunia telekomunikasi mengalami kemajuan yang sangat pesat, sehingga pengendalian berbasis suara telah dimanfaatkan dalam berbagai kebutuhan, khususnya dalam pengoperasian energi pada *smart building*.

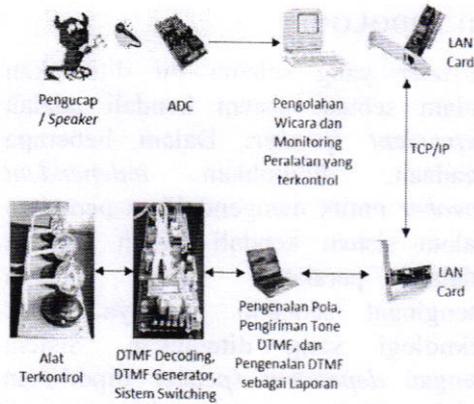
Indikasi sumber suara dibedakan berdasarkan pemilik system (*dependent*) dan lainnya (*independent*). Signifikasi suara berbasis *dependent* dan *independent* perlu diterjemahkan secara akurat sehingga dapat dikenali oleh perangkat kendalinya.

## METODOLOGI

*Speaker* yang selama ini diterapkan dalam sebuah sistem kendali adalah *dependent speaker*. Dalam beberapa keadaan, dibutuhkan *independent speaker* untuk mengendalikan peralatan dalam sistem kendali. Salah satunya adalah peralatan rumah tangga mengingat semakin majunya aspek teknologi yang diterapkan. Sistem dengan *dependent speaker* diperlukan salah satunya untuk sistem yang digunakan untuk kepetingan umum/bersama. Dengan adanya perbandingan *dependent* dan *independent speaker* maka dapat diketahui kehandalan sistem yang bisa mengendalikan dan memonitoring peralatan tertentu.

Sistem keseluruhan seperti yang tersaji pada Gambar 1. *Speaker* pada blok diagram Gambar 1 memberikan suaranya sebagai masukan lewat mikrofon. Masukan ini akan diolah dalam *PC client* mulai dari proses sampling, frame, *windowing*, FFT, LPC. Kemudian hasil dari LPC dicari pemadanan dengan LPC dari sinyal standar yang telah tersimpan sebagai *database* di dalam *PC client*. Hasil dari DTW tersebut akan dikirimkan ke *PC server* melalui TCP/IP untuk diproses menjadi suatu perintah ke perangkat keras yang dikendalikan. Perintah tersebut merupakan bentuk implementasi dari perintah yang diberikan dari *PC client*.

Sebaliknya dengan adanya perubahan kondisi pada peralatan yang diinginkan, maka terjadisinyal balikan yang berupa sinyal DTMF yang dikirimkan oleh perangkat keras ke *PC server*. Oleh *PC server* sinyal ini diolah untuk dikenali sebagai suatu kondisi tertentu yang akan dikirimkan ke *PC client* sebagai laporan tentang kondisi alat yang diinginkan.



**Gambar 1. Blok Diagram Sistem Pengendalian Lampu berbasis Wicara**

**Sinyal Suara Manusia**

Manusia menggunakan suara sebagai sumber informasi untuk mengkomunikasikan keinginan, ide dan perasaannya kepada orang lain. Sinyal wicara adalah sinyal yang dihasilkan oleh suara manusia yang mempunyai frekuensi kerja antara 0 sampai dengan 5000 Hz. Pada program *Snack* yang akan digunakan, frekuensi *sampling* diset pada 12 kHz. Frekuensi *sampling* ini dipilih karena biasa digunakan untuk *voice*, sedangkan untuk *telephone* biasa digunakan frekuensi *sampling* sebesar 8 kHz.

Gambar 2 merupakan proses ekstraksi suara yang akan digunakan dalam proses pengenalan wicara yang digunakan dalam sistem kendali. Proses ini dimulai dari masukan berupa *sample* suara yang akan digunakan dalam sistem.

Data yang digunakan sebagai data standar merupakan data 8 kombinasi kata, yaitu: satu nyala, satu mati, dua nyala, dua mati, tiga nyala, tiga mati, empat nyala, empat mati.



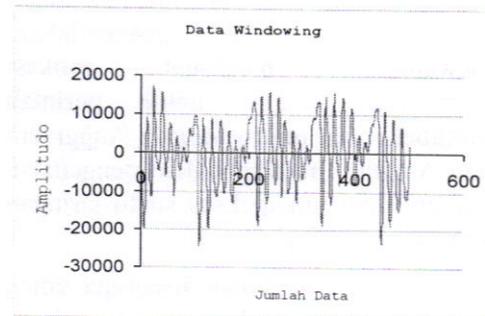
**Gambar 2. Proses Ekstraksi Sinyal Wicara**

**Sampling**

Sinyal wicara merupakan sinyal yang tidak terbatas dalam domain waktu (*finite time interval*). Untuk keperluan pemrosesan dalam transformasi *Fourier*, maka sinyal wicara harus dibentuk dalam potongan-potongan waktu yang terbatas (*infinite time interval*). Karena itu sinyal yang ada dipotong-potong dalam slot-slot interval waktu tertentu.

**Windowing**

*Windowing* merupakan proses pengurangan efek diskontinuitas pada awal dan akhir masing-masing sinyal yang telah di-*frame*<sup>[4]</sup>. Jenis *window* yang digunakan adalah *window Hamming*.

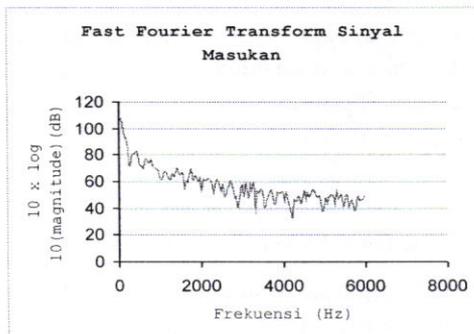


**Gambar 3. Hasil Windowing Data Sampel Kata “empatmati” milidetik ke-13900**

Gambar 3 menunjukkan proses *windowing* yang telah dilakukan. Dari gambar tersebut bisa diamati bahwa *window Hamming* menyebabkan sinyal yang disampel menjadi lebih halus bila dibandingkan dengan sinyal yang belum melalui proses *windowing*. Hal ini menunjukkan bahwa *windowing* berfungsi untuk mengurangi efek diskontinuitas pada ujung-ujung *frame*.

**Fourier Transform**

*Fast Fourier Transform* (FFT) adalah suatu metode yang sangat efisien untuk menyelesaikan Transformasi Fourier diskrit (DFT) yang banyak dipakai untuk keperluan analisa sinyal. Sehingga dengan DFT sinyal yang di-*sampling* dalam domain waktu ditransformasikan ke domain frekuensi.



**Gambar 4. Fast Fourier Transform Sinyal Masukan dengan Menggunakan Perangkat Lunak Tcl/Tk dan Snack**

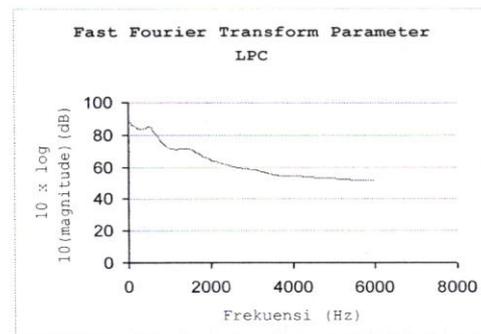
Setiap sinyal yang berasal dari alam merupakan sinyal analog yang bila diolah harus diubah ke dalam bentuk sinyal digital. Pengolahan dalam bentuk digital merupakan pengolahan dalam bentuk diskrit. Dalam penelitian ini, sinyal dalam domain waktu diubah ke dalam domain frekuensi dengan 256 titik. Karena hasil yang diperoleh berupa hasil fungsi konvolusi, maka hanya diambil 128 titik yang akan diolah dalam proses berikutnya. Sedangkan 128 sisanya tidak dipergunakan karena berupa pencerminan.

Gambar 4 merupakan sinyal suara yang sebelumnya berada dalam domain waktu telah diubah dalam domain frekuensi. Hasil *Fast Fourier Transform*, sinyal masukan menunjukkan sebaran spektrum frekuensi dari sinyal suara yang telah dihasilkan oleh pengucap antara 0 sampai dengan 6000 Hz.

**Linear Prediction Code**

*Linear Prediction Code* (LPC) merupakan proses pencarian ciri dari sinyal bicara yang akan diolah. *Linear Prediction Code* (LPC) secara umum sebagai pemroses *front-end* untuk pengenalan bicara, perlu diketahui alasan mengapa LPC banyak digunakan<sup>[5]</sup>, yaitu:

- LPC menyuguhkan model yang bagus untuk sinyal bicara. Cara yang digunakan LPC untuk analisa sinyal bicara sangat beralasan untuk pemisahan track dari sumber vokal.
- LPC adalah sebuah model traktabel secara analitis.
- Model LPC bekerja dengan baik pada aplikasi pengenalan.



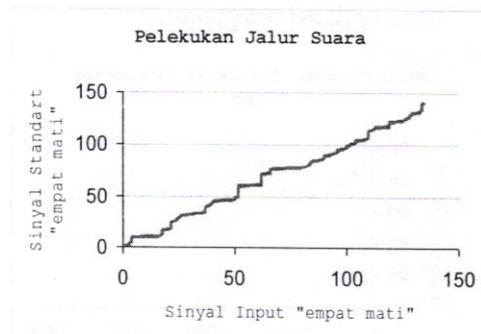
**Gambar 5. Fast Fourier Transform LPC dengan Menggunakan Perangkat Lunak TCL/Tk dan Snack**

Parameter LPC yang di-*fourier transform*-kan seperti ditunjukkan pada Gambar 5 akan menjadi selubung spektrum dari spektrum sinyal masukan. Untuk menggambarkan ciri atau fitur bicara dalam domain waktu, maka LPC spectrum di-*invers fast fourier*-kan.

**Dynamic Time Warping**

*Dynamic Time Warping* (DTW) merupakan cara untuk memadankan sinyal standar dengan sinyal input. Dengan menggunakan DTW, dua buah sinyal suara dibandingkan vektor-vektor

spektralnya seperti ditunjukkan pada Gambar 6, dimana terdapat proses perbandingan panjang sinyal antara dua buah sinyal tersebut, sehingga terjadi proses pemadanan panjang sinyal dengan cara memperpanjang dan menekuk vektor spektralnya. Memperpanjang dalam arti memperpanjang sinyal yang lebih pendek dengan memperhatikan perbandingan vektor spektralnya. Demikian pula dengan menekuk, yaitu menekuk sinyal yang lebih panjang dengan memperhatikan vektor spektralnya. Dengan panjang yang sama sebagai hasil DTW maka jarak euclidian antara sinyal dapat diperbandingkan atau dihitung.



Gambar 6. Pelekukan Jalur Suara dengan *Dynamic Time Warping*

### Socket Programming

*Socket programming* adalah suatu aplikasi program yang dapat membantu sistem komunikasi antar satu sistem komputer dengan sistem komputer lain dengan menggunakan protokol TCP/IP dan protokol lain yang terhubung.

*Speaker* memberikan suaranya sebagai masukan lewat mikrofon. Masukan ini akan diolah dalam *PC client* mulai dari proses *sampling, frame, windowing, FFT, LPC*. Kemudian hasil dari LPC dicari pemadanan (DTW) dengan LPC dari sinyal standar yang telah tersimpan sebagai basis data di dalam *PC client*. Hasil dari DTW tersebut akan dikirimkan ke *PC server* melalui TCP/IP

untuk diproses menjadi suatu perintah ke perangkat keras. Perintah tersebut merupakan bentuk implementasi dari perintah yang diberikan dari *PC client*.

Sebaliknya, dengan adanya perubahan kondisi pada peralatan yang diinginkan, maka terjadi sinyal balikan yang berupa sinyal DTMF yang dikirimkan oleh perangkat keras ke *PC server*. Oleh *PC server* sinyal DTMF ini diolah untuk dikenali sebagai suatu kondisi tertentu yang akan dikirimkan ke *PC client* sebagai laporan tentang kondisi alat yang diinginkan

### HASIL DAN PEMBAHASAN

Pada pengujian sistem, diambil sampel data dari dua orang yang terdiri dari dua orang pria. Pria pertama merupakan standar sekaligus pengisi suara untuk perintah yang digunakan. Pria kedua merupakan pembanding untuk mengetahui unjuk kerja sistem bila perintah dilakukan oleh suara yang bukan merupakan pengisi suara standar (*independent speaker*).

Identifikasi wicara sebagai perintah pengendalian lampu (Gambar 1) disimulasikan melalui suara pria 1 (*dependent speaker*) dan pria 2 sebagai *independent speaker*. Hasil pengujian dapat dilihat pada Tabel 1. *Speaker* memberikan perintah "satu nyala, satu mati, dua nyala, dua mati, tiga nyala, tiga mati, empat nyala, empat mati" kepada sistem yang diuji.

Pada Tabel 1 terlihat bahwa pengujian pada pria 1 yang digunakan sebagai standar (*dependent speaker*) memperoleh persentase keberhasilan rata-rata untuk semua perintah adalah sebesar 100%. Pengujian yang kedua adalah pengujian yang dilakukan oleh pria 2 atau *independent speaker* (pria lain yang suaranya tidak dipakai sebagai standar). Persentase keberhasilan rata-rata yang didapat untuk semua perintah

71,88%. Hal ini menunjukkan bahwa untuk suara yang dijadikan sebagai standar, sistem berhasil mengenali dan memberi respon sebagaimana perintah yang diberikan. Untuk suara yang tidak dijadikan sebagai standar keberhasilan sistem tidak setinggi jika dibandingkan dengan suara yang dijadikan standar. Hal ini menunjukkan sinyal masukan dan sinyal standar yang berbeda. Dari hasil pengujian menunjukkan ciri sinyal suara yang dihasilkan oleh *speaker* pertama sama dengan pemberi standar, sedangkan untuk *speaker* ke dua, *speaker* dan pemberi standar mempunyai ciri sinyal suara yang mendekati sama.

**Tabel 1. Tingkat Keberhasilan Sistem Perangkat Lunak**

NO.	PENGUCAP	PERINTAH	BANYAK PERCOBAAN	PERSENTASE KEBERHASILAN
1	Pria 1	1 ngala	8 kali	100%
2		1 mati	8 kali	100%
3		2 ngala	8 kali	100%
4		2 mati	8 kali	100%
5		3 ngala	8 kali	100%
6		3 mati	8 kali	100%
7		4 ngala	8 kali	100%
8		4 mati	8 kali	100%
RATA-RATA				100%
9	Pria 2	1 ngala	4 kali	100%
10		1 mati	4 kali	75%
11		2 ngala	4 kali	75%
12		2 mati	4 kali	50%
13		3 ngala	4 kali	50%
14		3 mati	4 kali	75%
15		4 ngala	4 kali	50%
16		4 mati	4 kali	100%
RATA-RATA				71,88%

**KESIMPULAN**

Sistem yang diuji mempunyai tingkat akurasi yang tinggi untuk pengucap masukan dan pengucap standar yang sama (*dependent speaker*). Hal ini dibuktikan dengan hasil pengujian yang dilakukan oleh pria pertama. Sedangkan untuk pengujian yang dilakukan oleh suara lain (*independent speaker*) tingkat

keberhasilannya bervariasi dari 50% sampai dengan 100%. Hal ini disebabkan karena suara pria kedua tidak pernah berinteraksi dengan sistem. Keberhasilan sistem ditunjukkan dengan ketepatan penyalaan/pemadaman lampu sesuai dengan yang diperintahkan melalui perintah suara.

**DAFTAR PUSTAKA**

1. Anggraeni, Triana Dewi, Astutik, Debi Puji, Pengenalan Wicara dengan Penutur Tak Bebas, PENS ITS 2003
2. Dewantara, Bima Sena Bayu, Aplikasi Pengenalan Wicara untuk Perintah Nirkabel Robot Mikro Mouse
3. Dutono, Titon, Penataran Kebahasaan Tahap I :Fonetik Dan Fonologi –Pengolahan Sinyal Wicara Digital, 2001
4. Gabel, Robert A, Richard A Robert, Sinyal dan Sistem Linier, Erlangga 1996
5. Rabiner, Lawrence, Huang Juang, Biing, Fundamentals of Speech Recognition, Prentice Hall International Inc, 1993
6. Proakis, John G, Malonakis, Dimitris G, Digital Signal Processing Principles, Algorithms, and Applications, Prentice Hall, Inc., New Jersey, 1995