

# ***SPEAKER IDENTIFICATION MENGGUNAKAN TRANSFORMASI WAVELET DISKRIT DAN JARINGAN SARAF TIRUAN BACK-PROPAGATION***

**Anny Tandyo<sup>1</sup>; Martono<sup>2</sup>; Adi Widyatmoko<sup>3</sup>**

**<sup>1, 2, 3</sup>Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Nusantara,  
Jln. K.H. Syahdan No.9, Palmerah, Jakarta Barat 11480**

## **ABSTRACT**

*Article discussed a speaker identification system. Which was a part of speaker recognition. The system identified a subject based on the voice from a group of pattern had been saved before. This system used a wavelet discrete transformation as a feature extraction method and an artificial neural network of back-propagation as a classification method. The voice input was processed by the wavelet discrete transformation in order to obtain signal coefficient of low frequency as a decomposition result which kept voice characteristic of everyone. The coefficient then was classified artificial neural network of back-propagation. A system trial was conducted by collecting voice samples directly by using 225 microphones in non soundproof rooms; contained of 15 subjects (persons) and each of them had 15 voice samples. The 10 samples were used as a training voice and 5 others as a testing voice. Identification accuracy rate reached 84 percent. The testing was also done on the subjects who pronounced same words. It can be concluded that, the similar selection of words by different subjects has no influence on the accuracy rate produced by system.*

**Keywords:** *speaker identification, wavelet discrete transformation, artificial neural network, back-propagation.*

## **ABSTRAK**

*Artikel membahas sistem speaker. Speaker identification merupakan bagian dari speaker recognition, yaitu sistem yang mengenali seorang subjek berdasarkan suara dari sekumpulan pola yang sudah disimpan sebelumnya. Sistem ini menggunakan transformasi wavelet diskrit sebagai metode ekstraksi fitur dan jaringan saraf tiruan back-propagation sebagai metode klasifikasi. Input suara diproses melalui transformasi wavelet diskrit untuk mendapatkan koefisien sinyal frekuensi rendah hasil dekomposisi yang menyimpan karakteristik suara setiap orang. Koefisien tersebut kemudian akan diklasifikasikan menggunakan jaringan saraf tiruan back-propagation. Pengujian sistem dilakukan dengan mengambil sampel suara secara langsung yang menggunakan mikrofon di ruangan yang tidak kedap suara sebanyak 225 buah; terdiri dari 15 subjek (orang) dan masing-masing memiliki 15 buah sampel suara. Sepuluh sampel digunakan sebagai suara pelatihan dan 5 sisanya digunakan untuk pengujian suara. Tingkat akurasi pengenalan mencapai 84 persen. Pengujian juga dilakukan pada subjek yang mengucapkan kata yang sama. Disimpulkan, pemilihan kata yang sama oleh subjek yang berbeda tidak berpengaruh terhadap tingkat akurasi yang dihasilkan oleh sistem.*

**Kata Kunci:** *speaker identification, transformasi wavelet diskrit, jaringan saraf tiruan, back-propagation.*

---

## **PENDAHULUAN**

### **Latar Belakang**

*Speaker recognition* adalah salah satu bidang pengenalan pola yang berkaitan dengan pemrosesan sinyal suara. Berbeda dengan *speech recognition* yang mengenali kata atau kalimat apa yang diucapkan, *speaker recognition* mengenali siapa pembicara yang mengucapkan kata atau kalimat tersebut. Dalam *speaker recognition*, data penting yang terkandung dalam sinyal suara diekstrak untuk kemudian diolah menjadi informasi yang dapat digunakan untuk proses lebih lanjut. Salah satu cabang dari bidang *speaker recognition* adalah *speaker identification*.

Sebuah sistem *speaker identification* mampu mengenali pemilik suara dari kumpulan data karakteristik suara yang sebelumnya telah disimpan. Berbagai teknik *speaker identification* berkembang lantaran meningkatnya kebutuhan identifikasi biometrik yang dapat bekerja efisien dan memiliki tingkat akurasi tinggi.

Sebuah sistem *speaker identification* yang baik

harus dapat mengambil fitur-fitur yang menjadi ciri khas suara seseorang, kemudian memproses fitur-fitur tersebut agar dapat diklasifikasikan ke dalam kelompok-kelompok tertentu untuk kemudian dikenali. Dalam beberapa penelitian mengenai pengenalan pola lainnya seperti pengenalan citra wajah, salah satu kombinasi metode pengambilan fitur dan klasifikasi yang sudah teruji adalah transformasi *wavelet* diskrit dan jaringan saraf tiruan. Dalam hubungannya dengan sinyal suara, transformasi *wavelet* diskrit dapat memisahkan sebuah sinyal menjadi sinyal berfrekuensi tinggi dan rendah untuk mendapatkan informasi yang menjadi ciri khas suara seseorang. Sementara itu, jaringan saraf tiruan memiliki kemampuan untuk mempelajari beragam karakteristik dalam suara seseorang. Dari permasalahan di atas, timbul gagasan untuk membuat sebuah sistem *speaker identification* yang menggunakan model *wavelet* sebagai ekstraksi fitur dan jaringan saraf tiruan sebagai metode klasifikasi.

Batasan permasalahan pada penelitian ini adalah: data suara yang dimasukkan berupa kata, frase, atau kalimat yang lama pengucapannya dibatasi hanya satu detik, dengan *sampling rate* 8000 kbps. Jika data suara lebih dari satu detik, data suara harus dimasukkan ulang; data suara yang

dimasukkan untuk dilatih dan dievaluasi berasal dari *file audio* atau hasil tangkapan menggunakan mikrofon; data suara dianalisis dan dikenali berdasarkan tingkat kemiripannya dengan data yang sudah dimasukkan sebelumnya dan tidak bisa digunakan untuk mengenali kata atau frase apa yang dimasukkan; metode ekstraksi fitur yang digunakan adalah transformasi *wavelet* diskrit dengan menggunakan fungsi *wavelet* induk "Daubechies-2"; metode klasifikasi yang digunakan adalah jaringan saraf tiruan dengan pendekatan *back-propagation*; serta program dikembangkan dengan menggunakan bahasa pemrograman "Delphi".

## Metode

Metode yang digunakan dalam penelitian ini meliputi 4 bagian pokok sebagai berikut. Pertama adalah metode analisis. Dalam penelitian ini, dilakukan beberapa cara untuk melakukan analisis; yakni dengan membaca literatur, buku, dan artikel yang berkaitan dengan *speaker identification*, jaringan saraf tiruan, serta teori-teori lain yang terkait. Kedua adalah metode perancangan. Dari hasil analisis dan studi pustaka yang diperoleh, dibuat sebuah rancangan sistem *speaker identification* yang dapat menerapkan transformasi *wavelet* dan jaringan saraf tiruan *back-propagation*. Ketiga adalah metode pengujian. Pengujian dilakukan terhadap kinerja *speaker identification*, sejauh mana program dapat mengenali suara seseorang dengan tepat. Pengujian dilakukan dengan mengambil sampel suara dari beberapa responden dengan menggunakan mikrofon. Keempat adalah metode evaluasi. Hasil pengujian yang diperoleh dianalisis lebih lanjut dan dibandingkan dengan *setting-an* konfigurasi lain, untuk ditarik kesimpulan mengenai kinerja sistem. Jika hasil evaluasi belum memadai, maka dilakukan pengujian tambahan.

## Tinjauan Pustaka

### Sinyal Suara

Suara adalah sinyal yang bergantung pada waktu (*time-dependent*). Oleh sebab itu, pengucapan kata yang sama dapat memiliki durasi yang berbeda. Pengucapan kata yang sama dengan durasi yang sama juga dapat memiliki perbedaan di bagian tengah, dikarenakan adanya perbedaan bagian dari kata yang diucapkan dengan kecepatan yang berbeda.

### Silence-Frame

*Silence-frame* merupakan *sample* pada suara yang tidak memiliki bunyi. Biasanya merupakan jeda antar kata, yang diucapkan ataupun kekosongan pada awal dan akhir dari sebuah pengucapan (Olsson, 2002: 27).

### Penguatan Suara

Menurut Mitra (1998: 4), penguatan suara atau amplifikasi suara adalah proses pengolahan suara meningkatkan nilai amplitudo dengan faktor pengali tertentu, sehingga amplitudo yang dihasilkan menjadi sejumlah kali lipat nilai semula.

### Normalisasi Audio

Normalisasi audio adalah proses pengolahan suara menaikkan atau menurunkan amplitudo atau volume dari sebuah *file* suara agar semua nilai *sample* di dalamnya berada pada rentang tertentu.

### Speaker Recognition

*Speaker recognition* adalah sistem yang dapat

mengenali seseorang dari suaranya (Sigmund, 2003: 7 - 18). Sistem ini mengekstrak fitur dari suara, memodelkannya, dan menggunakan model tersebut untuk membedakan seseorang berdasarkan suaranya. *Speaker recognition* sering disamakan dengan *speech recognition*, padahal keduanya memiliki definisi yang berbeda. *Speaker recognition* mengenali siapa yang berbicara, tetapi *speech recognition* mengenali apa yang diucapkan. *Speaker recognition* terdiri dari 2 tahap, yaitu pelatihan dan pengenalan (evaluasi).

### Transformasi Wavelet Diskrit

*Wavelet* adalah fungsi matematika yang memilah data menjadi berbagai komponen frekuensi, kemudian mempelajari masing-masing komponen dengan resolusi yang sesuai dengan faktor skalanya.

Dalam permasalahan diskrit, *filter* dari potongan frekuensi yang berbeda-beda digunakan untuk menganalisis sinyal pada skala yang berbeda. Sinyal *input* dilewatkan melalui sekelompok *high-pass filter* untuk menganalisis frekuensi tinggi, dan dilewatkan melalui sekelompok *low-pass filter* untuk menganalisis frekuensi rendah. Sinyal frekuensi rendah identik dengan informasi global yang terdapat pada sinyal *input*, sedangkan sinyal frekuensi tinggi identik dengan informasi detail dari sinyal *input*. Sinyal frekuensi rendah ini dapat dimanfaatkan untuk mengenali pola umum pada sinyal *input*.

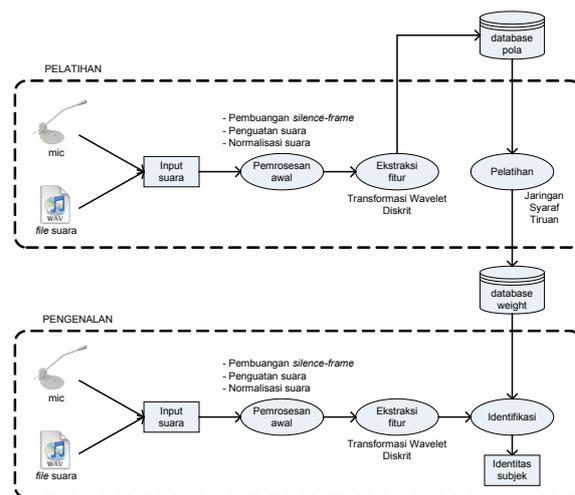
### Back-Propagation

*Back-propagation* merupakan suatu teknik untuk meminimalisasi gradien pada dimensi *weight* dalam jaringan saraf tiruan lapis banyak (Haykin, 1999: 202).

Algoritma pembelajaran *back-propagation* secara umum, yaitu: langkah pertama adalah inialisasi nilai *weight*, kedua adalah hitung nilai *output* pada *output layer*, ketiga adalah hitung perubahan pada *weight*, keempat adalah *update* nilai *weight*, dan kelima adalah hitung tingkat *error*.

## PEMBAHASAN

### Gambaran Umum



Gambar 1 Diagram Sistem Speaker Identification

Sistem *speaker identification* merupakan suatu sistem yang dapat mengenali pembicara atau orang yang mengucapkan kata-kata seperti halnya manusia, yang dapat mengenali suara seseorang. Manusia pada umumnya dapat mengenali suara seseorang dari warna suara atau ciri khas suara orang tersebut. Kemiripan suara orang tersebut dengan suara yang tersimpan dalam ingatan membuat orang dapat mengenali seseorang hanya dengan mendengar suaranya.

Pada sistem *speaker identification*, program mensimulasikan otak manusia dalam bentuk model jaringan saraf tiruan untuk menyimpan pola suara. *Input* suara didapat melalui *hardware* yang dapat menangkap suara yaitu mikrofon. Setiap subjek harus mengucapkan kata yang sama. Suara yang ditangkap kemudian diproses terlebih dahulu dengan membuang bagian tidak berisi suara yang tidak dibutuhkan (atau yang dikenal dengan *silence-frame*) dan melakukan penguatan atau amplifikasi pada suara. Setelah melalui pemrosesan dengan transformasi *wavelet* diskrit (*Discrete Wavelet Transform/DWT*), nilai koefisien hasil DWT disimpan ke dalam *database*, yang kemudian akan dilatih dalam sistem jaringan saraf tiruan. Setelah dilatih, sistem akan menyimpan nilai bobot atau *weight* yang nantinya akan digunakan pada tahap pengenalan (Gambar 1).

Proses identifikasi dilakukan dengan mengolah *input* suara, kemudian melakukan pengklasifikasian *input* dengan cara mencocokkan hasil proses dengan subjek yang paling sesuai berdasarkan hasil pelatihan sebelumnya.

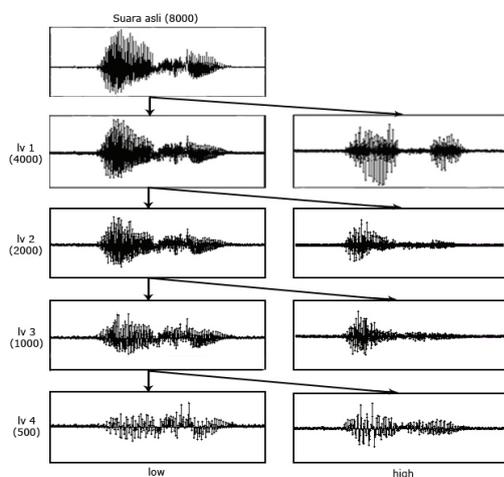
## Tahap Pemrosesan Awal

Pada tahap ini, *input* suara diproses terlebih dahulu agar siap untuk diproses lebih lanjut dengan menggunakan DWT. *Input* suara yang baru didapat memiliki *silence-frame* atau titik di mana tidak terdapat suara. Untuk mengatasi kelebihan *silence-frame* yang tidak berguna, pembuangan *silence-frame* harus dilakukan terlebih dahulu agar suara yang akan diproses dengan DWT sudah berupa suara yang bernilai.

Proses lain yang akan dilakukan adalah normalisasi suara dan atau penguatan suara. Normalisasi suara dilakukan agar semua *input* suara memiliki rentang nilai yang sama. Penguatan suara dilakukan untuk meningkatkan nilai suara yang rendah. Nilai rendah pada umumnya disebabkan oleh *volume* suara yang rendah pada saat pengambilan suara dilakukan.

## Transformasi Wavelet Diskrit

Transformasi *wavelet* diskrit dilakukan dengan mengaplikasikan *hi-pass filter* untuk mendapatkan sinyal frekuensi tinggi dan *lo-pass filter* untuk mendapatkan sinyal frekuensi rendah. Sinyal suara merupakan sinyal 1 dimensi, oleh sebab itu *filter hi-pass* dan *lo-pass* yang digunakan masing-masing sebanyak 1 *filter*. Proses transformasi ini disebut sebagai dekomposisi *wavelet* dan hasil dari dekomposisi *wavelet* disebut sebagai koefisien *wavelet*.



Gambar 2 Dekomposisi Wavelet

Sinyal suara asli yang ditransformasi *wavelet* untuk pertama kali (level 1) akan menghasilkan sebuah sinyal suara

frekuensi rendah dan sebuah sinyal suara frekuensi tinggi. Pada level berikutnya, transformasi *wavelet* diaplikasikan pada sinyal suara frekuensi rendah dari level sebelumnya.

Setiap kali dekomposisi *wavelet* (Gambar 2) akan menyebabkan panjang sinyal suara berkurang setengah dari panjang sinyal asli. Hasil dari dekomposisi *wavelet* adalah vektor fitur dari sinyal suara yang berupa sinyal frekuensi rendah. Vektor fitur inilah yang akan menjadi *input* bagi jaringan saraf tiruan.

Normalisasi dilakukan dengan cara membagi nilai koefisien dengan nilai maksimum yang dapat dicapai nilai koefisien *wavelet* pada level dekomposisi tersebut.

## Back-Propagation

Desain jaringan saraf tiruan *back-propagation* yang digunakan adalah sebagai berikut. Pertama, jumlah *hidden layer* yang digunakan adalah satu, dan fungsi aktivasi yang digunakan adalah fungsi sigmoid karena nilai keluaran yang diinginkan berada pada jangkauan 0 hingga 1.

## Strategi Representasi Data Output

Masing-masing *node* pada *output layer* mewakili sebuah kode identitas subjek. Untuk tahap pelatihan, nilai 0.98 digunakan sebagai *input* pada *node* di *output layer* yang mewakili kode identitas subjek tersebut, sedangkan *node* lainnya diberikan nilai 0.02. Untuk tahap identifikasi, kode identitas subjek ditentukan oleh *node* yang memiliki nilai tertinggi pada *output layer* (*winner take all*). Jumlah *node* yang dibutuhkan pada *output layer* adalah sebanyak jumlah subjek.

## Implementasi dan Evaluasi

### Spesifikasi Sistem

Aplikasi *speaker identification* ini menggunakan spesifikasi perangkat keras yang sama untuk semua pengujian, yaitu *Notebook Fujitsu N3510*, dengan spesifikasi: prosesor Intel Pentium Centrino M Processor 1,73 GHz, memori DDR2 512 MB, *harddisk* 60 GB, *graphic card* Ati Mobility Radeon X300, dan *sound card built-in* Realtek AC'97 Audio; mikrofon Genius; serta *speaker* Altec Lansing ATP 3.

### Prosedur Evaluasi

Pengujian pada sistem *speaker identification* ini menggunakan kumpulan data hasil perekaman melalui mikrofon. Setiap subjek dalam data memiliki 15 sampel suara dengan setiap sampel pada subjek yang sama memiliki kata sama yang diucapkan. 15 sampel suara tersebut kemudian dikelompokkan menjadi 2 kelompok, 10 sampel yang digunakan sebagai *input* untuk tahap pelatihan, dan 5 sampel yang digunakan sebagai *input* untuk tahap pengenalan. Pemisahan dimaksudkan agar sampel yang digunakan untuk pelatihan tidak digunakan kembali pada saat pengujian.

Pengujian yang dilakukan meliputi pengujian terhadap pengaruh level dan sinyal hasil dekomposisi *wavelet*, pengujian terhadap konfigurasi jaringan saraf tiruan *back-propagation*, dan pengujian terhadap pengaruh jumlah subjek dan jumlah pola tiap subjek yang dilatih.

### Evaluasi Pengaruh Penggunaan Frekuensi Sinyal Hasil Dekomposisi Wavelet, Penguatan Suara, dan Normalisasi

Pengujian ini bertujuan untuk mengevaluasi penggunaan sinyal berfrekuensi tinggi dan rendah hasil

dekomposisi *wavelet*, pengaruh penggunaan penguatan suara, dan normalisasi pada tahap pemrosesan awal. Hasilnya dapat dilihat pada Tabel 1 dan Gambar 3.

Dari hasil pengujian ini, solusi paling optimal terletak pada penggunaan sinyal frekuensi rendah hasil dekomposisi *wavelet* dan penggunaan normalisasi (tanpa amplifikasi) pada tahap pemrosesan awal.

### Evaluasi Pengaruh Level Dekomposisi *Wavelet*

Pengujian ini bertujuan untuk mengevaluasi pengaruh perubahan level dekomposisi *wavelet* yang digunakan pada saat ekstraksi fitur. Hasilnya dapat dilihat pada Tabel 2 dan Gambar 4.

Dari data hasil pengujian tersebut, hasil evaluasi paling optimal terdapat pada dekomposisi *wavelet* level 7.

### Evaluasi Pengaruh Jumlah *Node* pada *Hidden Layer*

Pengujian ini bertujuan untuk mengevaluasi pengaruh perubahan jumlah *node* pada *hidden layer* dan jumlah *node hidden layer* yang optimal. Hasilnya dapat dilihat pada Tabel 3 dan Gambar 5.

Dari hasil pengujian dengan jumlah *node* pada *hidden layer*, tingkat akurasi yang dihasilkan bervariasi dan mencapai titik tertinggi pada jumlah *node* sebanyak 62.

### Evaluasi Pengaruh Target *Error*

Pengujian ini bertujuan untuk mengevaluasi pengaruh perubahan target *error* yang digunakan pada jaringan saraf tiruan *back-propagation*. Hasilnya dapat dilihat pada Tabel 4 dan Gambar 6. Dari hasil pengujian ini, hasil terbaik dicapai pada target *error* 0,001.

### Evaluasi Pengaruh *Learning Rate*

Pengujian ini bertujuan untuk mengevaluasi pengaruh perubahan *learning rate* pada jaringan saraf tiruan *back-propagation*. Hasilnya dapat dilihat pada Tabel 5 dan Gambar 7. Dari data hasil pengujian tersebut, hasil optimal tercapai pada *learning rate* 0,1.

### Evaluasi Pengaruh Jumlah Data Pelatihan Per Subjek

Pengujian ini bertujuan untuk mengevaluasi pengaruh perubahan jumlah data pelatihan yang digunakan untuk setiap subjek. Hasilnya dapat dilihat pada Tabel 6 dan Gambar 8. Dari hasil pengujian tersebut, hasil paling optimal dari evaluasi ini terdapat pada jumlah data pelatihan sebanyak 10 sampel per subjek.

### Evaluasi Pengaruh Jumlah Subjek

Pengujian ini bertujuan untuk mengevaluasi pengaruh perubahan jumlah subjek yang ingin dikenali oleh aplikasi ini. Hasilnya dapat dilihat pada Tabel 7 dan Gambar 9. Dari hasil pengujian di atas, hasil paling optimal dicapai pada jumlah subjek sebanyak 10.

### Evaluasi Pengaruh Kata atau Frase yang Diucapkan

Pengujian ini dilakukan untuk menganalisis pengaruh kata atau frase yang diucapkan oleh subjek dalam proses pelatihan dan pengujian. Hasilnya dapat dilihat pada Tabel 8 dan Gambar 10. Dari hasil pengujian di atas, dapat dilihat bahwa tingkat akurasi pada subjek dengan kata yang sama dan dengan kata yang berbeda satu sama lain tidak mengalami perubahan yang signifikan.

Tabel 1 Data Hasil Pengujian Pengaruh Penggunaan Sinyal Hasil Dekomposisi *Wavelet*, Penguatan Suara, dan Normalisasi

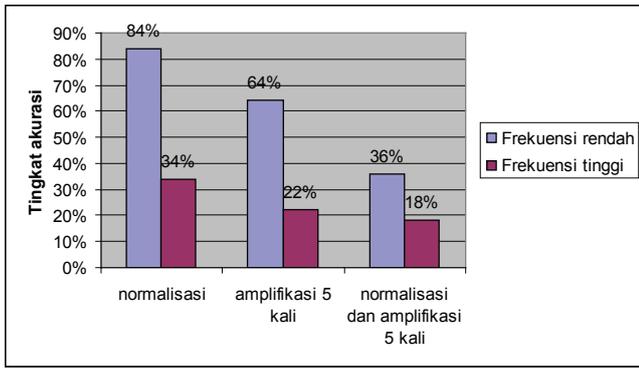
| No. | Frekuensi yang digunakan | Amplifikasi | Normalisasi | Tingkat akurasi |
|-----|--------------------------|-------------|-------------|-----------------|
| 1   | Rendah                   | Tidak       | Ya          | 84%             |
| 2   |                          | 5x          | Tidak       | 64%             |
| 3   |                          | 5x          | Ya          | 36%             |
| 4   | Tinggi                   | Tidak       | Ya          | 34%             |
| 5   |                          | 5x          | Tidak       | 22%             |
| 6   |                          | 5x          | Ya          | 18%             |

Tabel 2 Data Hasil Pengujian dengan Dekomposisi *Wavelet*

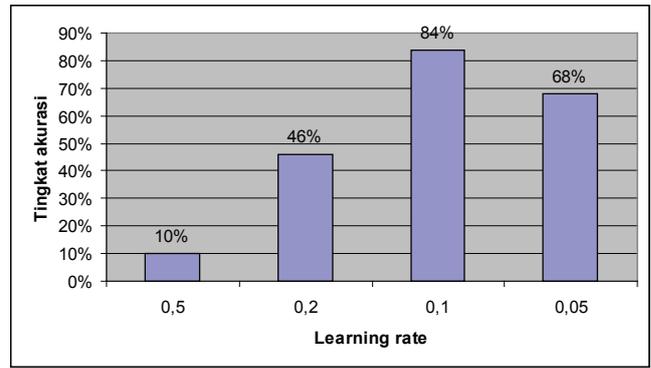
| No. | Level dekomposisi <i>wavelet</i> | Jumlah <i>node</i> DWT | Tingkat akurasi |
|-----|----------------------------------|------------------------|-----------------|
| 1   | 4                                | 500                    | 60%             |
| 2   | 5                                | 250                    | 66%             |
| 3   | 6                                | 125                    | 78%             |
| 4   | 7                                | 62                     | 84%             |
| 5   | 8                                | 31                     | 60%             |

Tabel 3 Tabel Ringkasan Pengujian Pengaruh Jumlah *Node* pada *Hidden Layer*

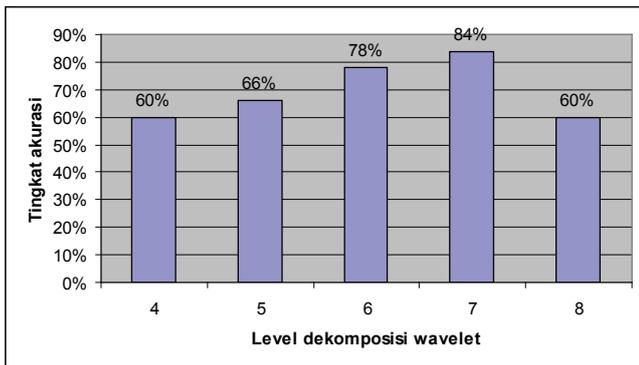
| No. | Jumlah <i>node</i> <i>hidden layer</i> | Tingkat akurasi |
|-----|--|-----------------|
| 1   | 50                                     | 52%             |
| 2   | 60                                     | 66%             |
| 3   | 62                                     | 84%             |
| 4   | 70                                     | 60%             |



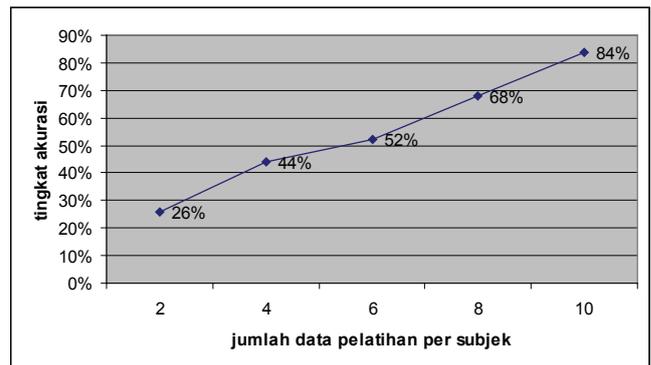
Gambar 3 Grafik Pengaruh Penggunaan Frekuensi Sinyal Hasil Dekomposisi *Wavelet*, Penguatan Suara, dan Normalisasi



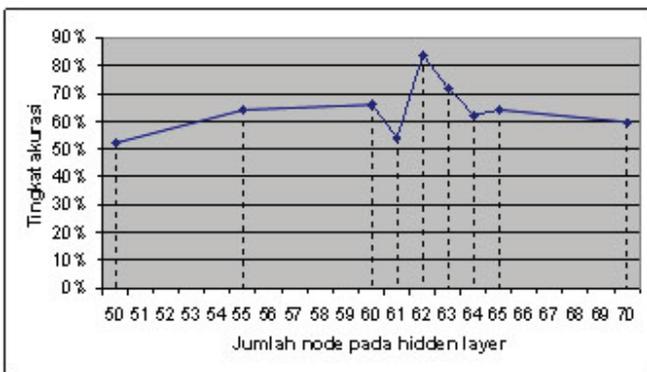
Gambar 7 Grafik Pengaruh *Learning Rate*



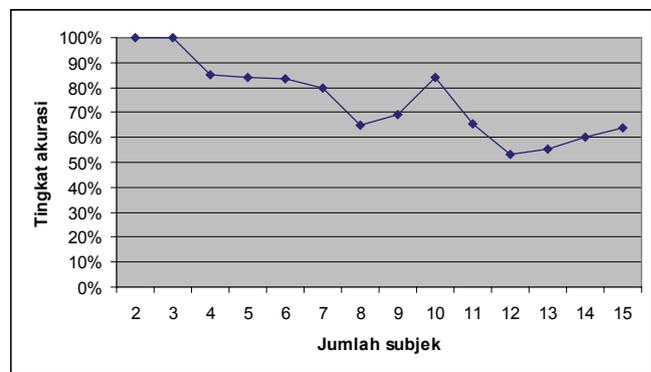
Gambar 4 Grafik Pengaruh Level Dekomposisi *Wavelet*



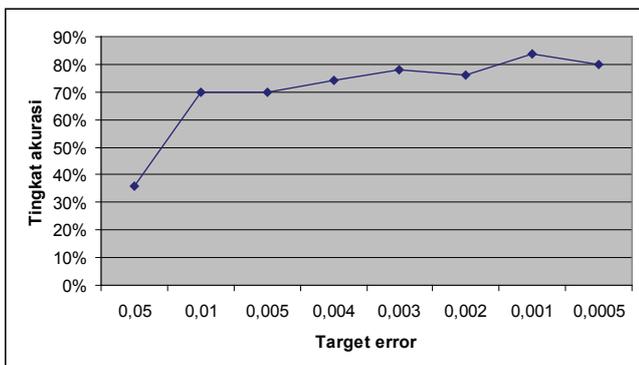
Gambar 8 Grafik Pengaruh Jumlah Data Pelatihan Per Subjek



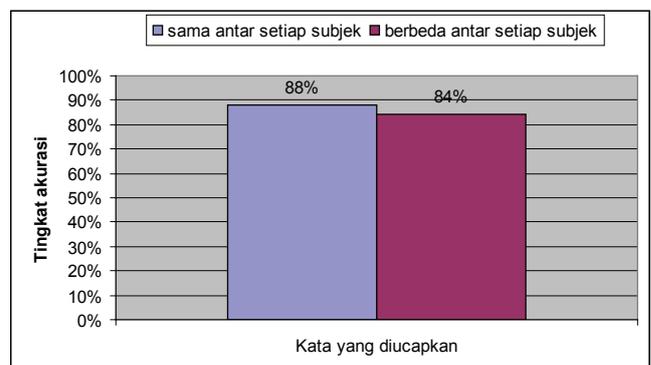
Gambar 5 Grafik Pengaruh Jumlah *Node* pada *Hidden Layer*



Gambar 9 Grafik Pengaruh Jumlah Subjek



Gambar 6 Grafik Pengaruh *Target Error*



Gambar 10 Grafik Pengaruh Kata atau Frase yang Diucapkan

Tabel 4 Tabel Ringkasan Hasil Pengujian Pengaruh Target Error

| No. | Target error | Tingkat akurasi |
|-----|--------------|-----------------|
| 1   | 0,05         | 36%             |
| 2   | 0,01         | 70%             |
| 3   | 0,005        | 70%             |
| 4   | 0,001        | 84%             |
| 5   | 0,0005       | 80%             |

Tabel 5 Tabel Ringkasan Hasil Pengujian Pengaruh Learning Rate

| No. | Learning rate | Tingkat akurasi |
|-----|---------------|-----------------|
| 1   | 0,5           | 10%             |
| 2   | 0,1           | 84%             |
| 3   | 0,05          | 68%             |

Tabel 6 Tabel Data Hasil Pengujian Pengaruh Jumlah Data Pelatihan Per Subjek

| No. | Jumlah data pelatihan per subjek | Tingkat akurasi |
|-----|----------------------------------|-----------------|
| 1   | 2                                | 26%             |
| 2   | 4                                | 44%             |
| 3   | 6                                | 52%             |
| 4   | 8                                | 68%             |
| 5   | 10                               | 84%             |

Tabel 7 Tabel Data Hasil Pengujian Pengaruh Jumlah Subjek

| No. | Jumlah subjek | Tingkat akurasi |
|-----|---------------|-----------------|
| 1   | 5             | 84%             |
| 2   | 10            | 84%             |
| 3   | 15            | 64%             |

Tabel 8 Tabel Evaluasi Pengaruh Kata atau Frase yang Diucapkan

| No. | Jumlah subjek | Kata yang diucapkan         | Tingkat akurasi |
|-----|---------------|-----------------------------|-----------------|
| 1   | 5             | sama antar setiap subjek    | 88%             |
| 2   | 5             | berbeda antar setiap subjek | 84%             |

## PENUTUP

Berdasarkan hasil evaluasi yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut. Pertama, kombinasi transformasi *wavelet* diskrit dan jaringan saraf tiruan *back-propagation* mampu menghasilkan suatu sistem *speaker identification* dengan tingkat akurasi yang cukup tinggi. Kedua, karakteristik suara dari setiap orang dapat diekstrak dengan menggunakan transformasi *wavelet* diskrit,

yang akan tersimpan dalam sinyal global berfrekuensi rendah hasil dekomposisi, dan digunakan dalam tahap pelatihan. Dengan sinyal global ini, sistem *speaker identification* dapat mengenali subjek dari kumpulan subjek yang mengucapkan kata atau frase yang sama. Ketiga, tahap pemrosesan awal yang dibutuhkan untuk menghasilkan tingkat akurasi pengenalan yang optimal adalah penggunaan pemotongan *silence-frame* dan normalisasi suara tanpa penguatan (amplifikasi). Keempat, penambahan level dekomposisi *wavelet* akan meningkatkan tingkat akurasi pengenalan sampai tingkat tertentu. Jika sesudahnya proses dekomposisi *wavelet* masih terus dilakukan, maka tingkat akurasi akan menurun karena informasi yang dihasilkan semakin sedikit dan umum. Kelima, transformasi *wavelet* diskrit mampu memperkecil pengaruh yang disebabkan oleh *noise* pada sinyal suara *input*. Hal ini disebabkan karena fitur yang digunakan untuk pengenalan adalah koefisien DWT sinyal global yang tidak peka terhadap *noise*. Keenam, jaringan saraf tiruan *back-propagation* akan bekerja dengan tingkat akurasi yang lebih baik, jika jumlah data pelatihan yang digunakan untuk tiap subjek semakin banyak dan kondisi suara *input* untuk tiap subjek bervariasi. Ketujuh, sistem *speaker identification* ini dapat diterapkan ke dalam sistem otentikasi, di mana jumlah subjek yang ingin dikenali tidak sering berubah.

Beberapa saran yang diusulkan untuk memperbaiki kinerja aplikasi adalah sebagai berikut. Pertama, menambahkan fitur *noise reduction* pada tahap pemrosesan awal untuk mengurangi *noise* pada *input* dan fitur pembuangan *silence-frame* yang terletak di tengah sinyal suara sebagai jeda pengucapan, agar tingkat akurasi semakin baik. Kedua, jika sistem *speaker identification* ingin diterapkan pada kondisi di mana jumlah subjek sangat banyak dan dapat berubah setiap saat, maka diperlukan modifikasi algoritma pembelajaran yang lebih sesuai mengingat waktu pelatihan *back-propagation* cukup lama, meskipun waktu eksekusinya sangat singkat seperti *incremental learning* atau pembobotan per subjek.

## DAFTAR PUSTAKA

- Allen, J. B. (1994). *How Do Humans Process and Recognize Speech*. IEEE Trans. on Speech and Signal Proc., Vol 2 (4), p. 567 - 576.
- Azcarraga, A. (1999). *Artificial Neural Networks*. <http://www.comp.nus.edu.sg/~pris/ArtificialNeuralNetworks/index.html>.
- Campbell, Joseph P. JR. (1997). *Speaker Recognition: A Tutorial*. Proceedings of the IEEE, Vol. 85 (9).
- Djuhana, Dede. (1998). *Bunyi*. <http://www.fisika.ui.ac.id/~dede/sound.pdf>.
- Fausett, Laurene. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. New Jersey: Prentice Hall.
- Graps, A. (1995). *An Introduction to Wavelets*. IEEE Computational Science and Engineering, Vol. 2 (2), p. 50 - 61.
- Haykin, Simon. (1999). *Neural Network: A Comprehensive Foundation*. New Jersey: Prentice Hall.
- Heisenberg, W. (1927). *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*. *Zeitschrift für Physik*, Vol. 43, p. 172 - 198.
- Jain, Anil K., Ruud Bolle, dan Sharath Pankanti. (2002). *Biometrics: Personal Identification in Networked Society*. New York: Kluwer Academic Publishers.
- Kamen, Edward W., dan Bonnie S. Heck. (2000). *Fundamentals of Signals and Systems Using the Web and Matlab*. Edisi kedua. New Jersey: Prentice Hall.

- Kulkarni, A. D. (2001). *Computer Vision and Fuzzy-Neural Systems*. New Jersey: Prentice Hall.
- McClelland, James L., dan David E. Rumelhart. (1989). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. Massachusetts: MIT Press.
- Mitra, S. K. (1998). *Digital Signal Processing: A Computer-Based Approach*. New York: McGraw-Hill.
- Olsson, Johan. (2002). *Text Dependent Speaker Verification with a Hybrid HMM/ANN System*. Stockholm: Kungliga Tekniska Högskolan.
- Polikar, R. (2001). *The Wavelet Tutorial*. <http://users.roman.edu/~polikar/WAVELETS/WTutorial.html>.
- Proakis, J. G., dan Manolakis, D. G. (1992). *Digital Signal Processing: Principles, Algorithms, and Applications*. Edisi kedua. New York: Macmillian Publishing Company.
- Rabiner, L., dan Juang, B.H. (1993). *Fundamentals of Speech Recognition*. New Jersey: Prentice-Hall.
- Reynolds, Douglas A., dan Larry P. Heck. (2000). *Automatic Speaker Recognition: Recent Progress, Current Applications, and Future Trends*. <http://www.ll.mit.edu/IST/pubs/aaas00-dar-pres.pdf>
- Russel, S., dan Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Edisi kedua. New Jersey: Pearson Education.
- Sakoe dan Chiba. (1978). *Dynamic Programming Algorithm Optimization for Spoken Word Recognition*. IEEE Trans. Acoustics, Speech, Signal Processing, ASSP Vol. 26 (1), p. 43 - 49.
- Shannon, C. E. (1949). *Communication in the Presence of Noise*. Proc. Institute of Radio Engineers, Vol. 37 (1), p. 10 - 21.
- Sigmund, Milan. (2003). *Voice Recognition by Computer*. Marburg: Tectum Verlag.
- Theodoridis, Sergios., dan Konsantinos Koutroumbas. (2006). *Pattern Recognition*. Edisi ketiga. Amsterdam: Elsevier.
- Wilson, Scott. (2003). *WAVE PCM Soundfile Format*. <http://ccrma.stanford.edu/CCRMA/Courses/422/projects/WaveFormat/>.