

IDENTIFIKASI PEMBICARA DENGAN WAVELET ORTHOGONAL COIFLET

(*Speaker Identifier with Wavelet Orthogonal Coiflet*)

KETUT AGUSTINI

Jurusan Diploma III Manajemen Informatika, FPTK IKIP Negeri Singaraja

eghee2006@Gmail.com, adisti_cecilia@yahoo.com

ABSTRACT

Transformasi Wavelet merupakan sarana yang mulai populer pada pemrosesan sinyal, seperti citra dan suara. Transformasi Wavelet merupakan alat yang sesuai untuk menganalisis sinyal nonstationer (seperti suara) dan memiliki kemampuan yang baik dalam melokalisasi frekuensi dan waktu sinyal.

Pada penelitian ini akan menggunakan tipe wavelet berbasis orthogonal yaitu coiflet berorde 4 dengan tingkat dekomposisi 10 dan 15. Pada pencocokan polanya menggunakan JST multi-layer perceptron.

Hasil eksperimen menunjukkan bahwa Transformasi Wavelet merupakan suatu metode ekstraksi ciri yang handal untuk identifikasi pembicara yang mampu bersaing dengan tools lainnya dan Wavelet tipe coiflet memberikan tingkat pengenalan sebesar 84%.

Keywords: *Transformasi Wavelet, Identifikasi Pembicara, Proses Sinyal, Wavelet basis orthogonal, Tingkat Dekomposisi, JST Multilayer Perceptron*

PENDAHULUAN

Perkembangan teknologi memungkinkan dilakukannya proses suara menggunakan komputer, baik untuk analisis suara (*speech analysis*), maupun sintesis suara (*speech syntesis*). Untuk keperluan analisis, sinyal suara analog mula-mula diubah menjadi sinyal digital, sehingga dapat diproses menggunakan komputer. Pengubahan dilakukan dengan cara mencuplik (*sampling*) dan mengkuantisasi contoh (*sampel*) sinyal suara dengan panjang segmen tertentu menggunakan ADC (*analog to digital converter*). Selanjutnya untuk sintesis, sinyal digital hasil pemrosesan diubah kembali menjadi sinyal analog menggunakan DAC (*digital to analog converter*).

Analisis suara memainkan peranan yang penting dalam pengenalan ucapan (*speech recognition*) atau pemahaman ucapan (*speech understanding*) dan identifikasi pembicara (*speaker Identification*) atau pemeriksaan pembicara (*speaker verification*).

Sistem identifikasi pembicara adalah satu cabang dari pengolahan sinyal suara. Sistem ini ditujukan untuk mengenali pembicara melalui input suara dari seorang pembicara. Identifikasi pembicara (*speaker Identification*) berbeda dari pengenalan ucapan (*speech recognition*), di mana pengenalan ucapan lebih dikonsentrasikan pada ekstraksi dari beberapa bagian informasi pesan yang didalamnya berisikan teks yang diucapkan. Teks yang diucapkan ini mengandung unit-unit linguistik yang dikenal sebagai fonem, kata-kata kunci suatu pesan ataupun pesan yang lengkap. Sedangkan identifikasi pembicara (*speaker identification*) menekankan pada analisis suara yang diucapkan oleh tiap pembicara sehingga dapat dibedakan karakteristik unik suara dari tiap pembicara tersebut. Banyak faktor yang mempengaruhi kesalahan aplikasi identifikasi pembicara menyangkut hal sebagai berikut (Campbell,1997), yaitu kesalahan membaca atau pengucapan frase yang telah

ditetapkan, tekanan emosional (*stress*) pengucapan, inkonsistensi ruang akustik (misalnya adanya *noise*), kesalahan *channel* transmisi (misalnya menggunakan mikrofon yang berbeda saat identifikasi dan verifikasi), kondisi pembicara yang mempengaruhi *vocal tract*-nya (flu atau *mood*), usia pembicara dan peniruan suara (*mimicry*).

Proses identifikasi dengan suara yang merupakan bagian dari sistem pengenalan pembicara memiliki keuntungan secara ekonomis dibandingkan dengan identifikasi secara biometrik lainnya seperti identifikasi pada wajah, sidik jari, tanda tangan, retina dan lain-lain. Identifikasi dengan suara hanya membutuhkan alat tambahan berupa mikrofon dan kartu suara sedangkan karakteristik lain misalnya sidik jari atau wajah membutuhkan alat tambahan seperti *scanner*. Hal ini sedikit banyak menekan biaya pengembangan sistem (Campbell,1997).

Sinyal suara manusia mempunyai tingkat variabilitas yang sangat tinggi. Suatu sinyal suara yang dikeluarkan oleh pembicara yang berbeda-beda menghasilkan pola ucapan yang berbeda-beda pula. Salah satu cara yang paling handal dalam pengenalan pola adalah Jaringan Syaraf Tiruan (JST). JST mampu menyelesaikan persoalan rumit yang sulit atau bahkan tidak mungkin jika diselesaikan dengan menggunakan komputasi konvensional. Salah satu jenis JST yang mampu memberikan unjuk kerja yang bagus adalah JST dengan arsitektur *Multi-Layer Perceptron* (MLP) dan pembelajaran *Backpropagation* (Fu,1994)

Penelitian mengenai identifikasi pembicara telah banyak dilakukan dengan berbagai metode pemrosesan sinyal, seperti Linier Prediction Coding (LPC), Mel Frequency Cepstrum Coefficients (MFCC), Neural Predictive Coding (NPC), dan sebagainya, yang mana keseluruhan metode diatas berbasiskan Transformasi Fourier, dan tingkat identifikasinya telah mencapai 100 %.

Untuk itu, akan dicobakan suatu konsep atau pendekatan lain tanpa berbasiskan Transformasi Fourier yaitu Transformasi Wavelet.

Transformasi Wavelet merupakan sarana yang mulai populer untuk pemrosesan sinyal, seperti citra dan suara, dan transformasi ini belum banyak diaplikasikan untuk analisis suara, khususnya untuk identifikasi pembicara menggunakan teks berbahasa Indonesia. Dalam praktek, Transformasi Wavelet digunakan untuk ekstraksi ciri dalam sistem pengenalan suara karena mempunyai karakter khusus yang sesuai untuk analisis sinyal, termasuk sinyal suara. Transformasi wavelet sinyal suara menghasilkan resolusi waktu yang baik pada frekuensi tinggi dalam menentukan lokasi awal suara dan parameterisasi ciri suara durasi pendek serta mampu menganalisis sinyal diskontinu (*non stationary*) secara akurat (Krisnan,1994).

Bertolak dari hal tersebut, seperti besarnya manfaat dan banyaknya faktor yang mempengaruhi hasil identifikasi pembicara, maka penelitian ini masih relevan dan layak dilakukan. Pada penelitian ini akan dikembangkan model sistem identifikasi pembicara dengan *text-dependent* berbahasa Indonesia menggunakan Transformasi Wavelet Diskret berbasis orthogonal yaitu *Coiflet* dan Jaringan Syaraf Tiruan *Multi Layer Perceptron* sebagai pengenalan pola.

Prinsip Dasar Identifikasi Pembicara

Pengenalan pembicara dapat diklasifikasikan ke dalam tiga tahap yaitu identifikasi, deteksi dan verifikasi. Identifikasi pembicara merupakan proses untuk menentukan identitas pembicara melalui suara yang telah diucapkan, sedangkan deteksi pembicara merupakan proses penemuan suara pembicara dari sekumpulan suara, dan verifikasi pembicara merupakan proses untuk memverifikasi kesesuaian suara pembicara dengan identitas yang diklaim oleh pembicara. Pengenalan pembicara lebih menitikberatkan pada pengenalan suara pembicara dan tidak pada pengenalan ucapan pembicara (Ho,1998) .

Metode identifikasi pembicara yang merupakan bagian dari pengenalan pembicara (**Gambar 1**), dapat dibagi ke dalam metode *text-independent* dan *text-dependent*. Pada sistem *text-independent*, model pembicara meng-*capture* karakteristik ucapan seseorang melalui sinyal ucapan dengan mengabaikan apa yang diucapkannya, dalam artian kata-kata yang diucapkan sembarang (bebas). Sebaliknya pada sistem *text-dependent*, pengenalan identitas pembicaranya didasarkan pada ucapan seseorang dengan kata-kata yang spesifik atau telah disepakati, seperti *password*, *card numbers*, kode PIN dan sebagainya (Mudry,1997).

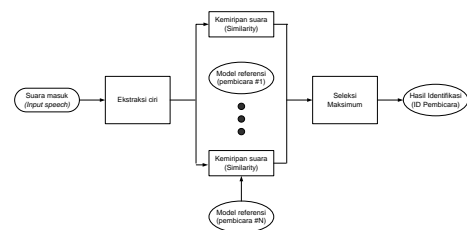


Gambar 1. Klasifikasi sistem pemrosesan sinyal suara

Semua sistem identifikasi pembicara melalui dua proses penting (seperti pada **Gambar 2**) yaitu *feature extraction* dan *feature matching*. *Feature extraction* merupakan proses mengekstraksi data hasil akuisisi sehingga dihasilkan data yang berdimensi lebih kecil, yang nantinya digunakan untuk merepresentasikan tiap-tiap pembicara. *Feature matching* menyangkut prosedur aktual yang mengidentifikasi pembicara yang tidak dikenal dan membandingkan fitur ekstraksi suara yang dimasukan dengan salah satu dari himpunan pembicara yang telah dikenal.

Sistem Identifikasi pembicara juga menyajikan dua sesi yang berbeda, yang pertama menunjukkan sesi pendaftaran (*enrollment sessions*) atau fase training, sedangkan yang kedua menunjukkan sesi operasi atau fase testing. Di dalam fase training, tiap pembicara yang telah terdaftar memasukkan contoh (*sampel*) suaranya sehingga sistem dapat mulai dibangun atau dilatih berdasarkan *reference* model pembicara tadi.

Berbagai Metode pemrosesan sinyal suara telah dilakukan para peneliti sampai tahun 2004, dalam sistem identifikasi pembicara, berikut adalah metode-metode yang telah diterapkan dengan tingkat identifikasi yang telah dicapainya, dituangkan pada **Tabel 1**.



Gambar 2. Struktur dasar dari sistem identifikasi pembicara.

Tabel 1. Metode-metode yang digunakan penelitian sebelumnya

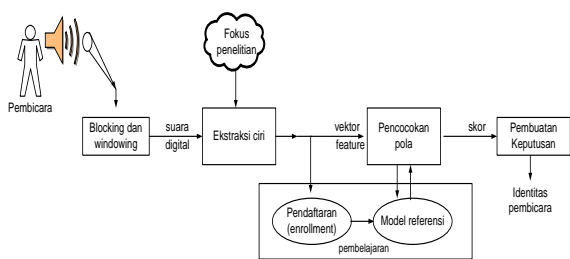
Metode pemrosesan sinyal suara	Tingkat identifikasi (%)
LPC (Linier Predictive Coding)	90.61
LPCC	96.73
MFCC (Mel Frequency Cepstrum Coefficient)	97.55
PLP	86.12
NPC (random initialization)	61.63
NPC (linier initialization)	100

Sumber : Universite Pierre&MarieCurie, LA Science A Paris,2004

Tahapan Sistem Identifikasi Pembicara

Secara umum sistem identifikasi pembicara mempunyai tahapan sebagai berikut dengan diagram bloknya diilustrasikan pada **Gambar 3** (Campbell,1997) :

- a) **Akuisisi data suara digital**, yaitu proses untuk mengakuisisi ucapan pembicara (dalam sinyal analog) dan mengubahnya menjadi sinyal digital. Sinyal digital yang terbentuk berupa suatu vektor yang merepresentasikan suara pembicara.
- b) **Frame blocking** dan **windowing**, yaitu *frame blocking* merupakan proses segmentasi sinyal suara digital yang telah diakuisisi ke dalam durasi tertentu, sedangkan *frame windowing* adalah proses yang bertujuan untuk meminimalkan diskontinuitas (*non-stationary*) sinyal pada bagian awal dan akhir sinyal suara.
- c) **Ekstraksi Ciri (*Feature extraction*)**, yaitu mengekstrak data hasil akuisisi sehingga dihasilkan data yang berdimensi lebih kecil tanpa merubah karakteristik sinyal suara.
- d) **Pembentukan model referensi pembicara**, merupakan tahapan pembelajaran dan akan membentuk suatu model referensi agar sistem dapat mengenali pembicara. Tahap ini memerlukan data berupa vektor-vektor ciri hasil dari ekstraksi ciri yang mencakup seluruh pembicara. Model referensi yang terbentuk akan digunakan dalam pencocokan pola. Pembentukan model referensi pembicara merupakan tahapan khusus yang dilakukan pada waktu awal sebelum sistem siap digunakan. Tahap ini hanya dilakukan sekali dan setelah dilakukan maka sistem siap untuk digunakan.
- e) **Pencocokan pola (*pattern matching*)**, yaitu proses pencocokan pola dengan menerima data yang telah diolah oleh ekstraksi ciri sebagai data input. Proses tersebut akan mencocokkan pola data masukan (*input*) dengan model referensi dan akan memberikan hasil berupa besarnya skor kesesuaian data input dengan pola-pola referensi yang ada.
- f) **Pembuatan keputusan**, Pembuatan keputusan akan menerima skor hasil pencocokan pola. Pada sistem identifikasi, pembuatan keputusan akan menentukan identitas pembicara .



Gambar 3 . Tahapan identifikasi pembicara.

Akuisisi Data Suara Digital

Sebagian besar sinyal-sinyal untuk maksud praktis, seperti suara, sinyal biologis, sinyal *seismic*, sinyal radar, sinyal sonar, dan berbagai sinyal komunikasi seperti

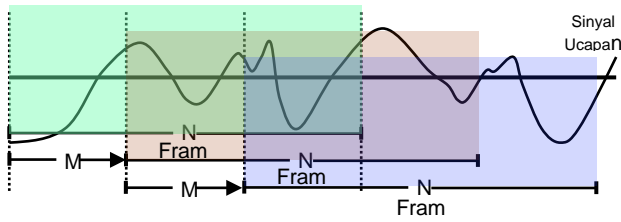
sinyal audio dan video, adalah sinyal analog (Proakis&Manolakis,1997). Untuk memproses sinyal analog dengan alat digital, pertama-tama perlu mengkonversinya menjadi bentuk digital yaitu mengkonversi menjadi suatu deret angka yang mempunyai presisi nilai terbatas. Prosedur ini dinamakan konversi *analog to digital* (A/D), dan alat yang sesuai dinamakan pengkonversi A/D (ADCS). Secara konsepsi, konversi A/D melalui 3 tahapan proses yaitu (1). Proses sampling (2). Kuantisasi dan (3). Pengkodean, (Proakis&Manolakis,1997)

Umumnya frekuensi sampling (*fs*) yang digunakan untuk analisis sinyal suara berkisar pada rentang 6-20 kHz (Mudry,1997). Jika dilakukan proses sampling selama *t* detik dengan *fs* adalah *f* (*fs=f*) maka akan diperoleh suatu vektor *V* yang mempunyai elemen sebanyak *f x t* sebagai berikut : $V^T = [V_1 V_2 V_3 \dots V_{ft}]$. Vektor *V* merepresentasikan sinyal suara dan dapat disebut sebagai sinyal suara digital.

Frame Blocking dan Windowing

Dalam interval waktu yang panjang, pola gelombang suara tidak *stationer*, tetapi dalam waktu yang cukup pendek (10-30 milidetik) dapat dikatakan *stationer*. Hal ini dikarenakan kecepatan perubahan spektrum suara berkaitan dengan kecepatan perubahan organ-organ penghasil suara pada manusia dan hal ini dibatasi oleh keterbatasan fisiologi (Xafopoulos,2001). Berdasarkan hal tersebut, sinyal suara digital yang telah diakuisisi dapat dibagi-bagi menjadi segmen-segmen dengan durasi 10-30 milidetik. Segmen sinyal suara digital ini disebut dengan *frame* dan proses pembentukan frame-frame disebut dengan *frame blocking* dengan tiap *frame* direpresentasikan dalam sebuah vektor.

Setiap *frame* akan memiliki jumlah sampel yang sama, misalnya *N* sampel. *Frame* kedua adalah *frame* yang juga memiliki *N* sampel yang posisi awal *frame*nya bergeser sebanyak *M* sampel dari posisi awal *frame* pertama. Begitu pula *frame* ketiga, dengan *N* sampel yang posisi awal *frame*nya bergeser sebanyak *M* sampel dari posisi awal *frame* kedua atau sebanyak *2M* sampel dari posisi awal *frame* pertama. Demikian pula seterusnya hingga *frame* terakhir *M* dapat diperoleh dari $M=(1/3)N$ atau $M = (a/b)N$ dimana *a* dan *b* adalah bilangan asli, $a \leq b$ dan $M \leq N$. *Overlap* antara satu *frame* dengan *frame* sebelumnya adalah *N – M* sampel (Rabiner dan Juang, 1993). Jika ternyata dalam *frame* terakhir jumlahnya kurang dari *N* maka *frame* terakhir tersebut dapat diabaikan karena umumnya sinyal suara pada bagian akhir tidak mengandung informasi yang penting. Ilustrasi tentang pembentukan *frame* dapat dilihat pada **Gambar 4**.



Gambar 4. Pembentukan frame pada sinyal suara .

Tahap selanjutnya dari pemrosesan sinyal adalah membuat *window* terhadap masing-masing frame dengan tujuan untuk meminimalkan ketidakkontinuan pada awal dan akhir setiap frame. Umumnya, *window* yang digunakan adalah window Hamming. Pembentukan window Hamming menggunakan formula:

$$w[k+1] = 0.54 - 0.46 \cos\left(2\pi \frac{k}{n-1}\right); \quad k=0, \dots, n-1$$

dimana n adalah lebar frame. Setelah frame *blocking* dan frame *windowing* selesai dilakukan maka pada tiap-tiap frame dilakukan ekstraksi ciri (*feature extraction*).

Ekstraksi Ciri (Feature Extraction)

Tujuan utama dari ekstraksi ciri adalah untuk mereduksi ukuran data tanpa merubah karakteristik dari sinyal suara dalam setiap framenya. Cara kerjanya adalah dengan mengkonversikan bentuk sinyal suara ke dalam bentuk representasi secara parameter. Banyak metode yang ada untuk ekstraksi ciri diantaranya Linier Predictive Coding (LPC), Mel-Cepstrum, dan *filter bank*. Transformasi Wavelet Diskret (TWD) adalah bagian dari *filter bank* selain Discret Fourier Transformasi (DFT), merupakan metode baru yang mempunyai karakteristik khusus yang sesuai untuk analisis sinyal, termasuk sinyal suara. Transformasi ini dapat digunakan untuk memberikan informasi mengenai sinyal secara bersamaan dalam domain waktu dan frekuensi.

Transformasi Wavelet Diskret

Transformasi adalah proses merepresentasikan suatu sinyal ke dalam *domain* / kawasan lain. Tujuan dari transformasi adalah untuk lebih menonjolkan sifat atau karakteristik sinyal tersebut.

Definisi wavelet (secara harfiah berarti “gelombang kecil”) adalah himpunan fungsi dalam ruang vektor $L^2\mathbf{I}$, yang mempunyai sifat-sifat sebagai berikut (Burrus *et al*,1998):

- berenergi terbatas
- merupakan fungsi *band-pass* pada domain frekuensi
- merupakan hasil penggeseran (*translasi*) dan penskala (*dilatasi*) dari sebuah fungsi tunggal (induk), yaitu

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$$

dengan $a, b \in \mathbf{R}$ (bilangan nyata), dan $a \neq 0$. Dalam hal ini a adalah parameter penskala dan b adalah parameter

penggeser posisi terhadap sumbu t . Faktor normalisasi $|a|^{-1/2}$ digunakan untuk memastikan bahwa $\|\psi_{a,b}(t)\| = \|\psi(t)\|$.

Pada dasarnya, Transformasi Wavelet merupakan sebuah teknik pemrosesan sinyal *multiresolusi*. Dengan sifat penskalaannya, wavelet dapat memilah-milah suatu sinyal data berdasarkan komponen frekuensi yang berbeda-beda. Dengan demikian tiap-tiap bagian dapat dipelajari berdasarkan skala resolusi yang sesuai, sehingga diperoleh gambaran data secara keseluruhan dan *detail*.

Jadi teori wavelet didasari oleh pembangkitan sejumlah tapis (*filter*) dengan cara menggeser dan menskala suatu wavelet induk (*mother wavelet*) berupa tapis pelewat tengah (*band-pass filter*). Dengan demikian hanya diperlukan pembangkitan sebuah tapis. Tapis lain mengikuti aturan penskalaan, baik pada kawasan waktu maupun kawasan frekuensi. Penambahan skala wavelet akan meningkatkan durasi waktu, mengurangi lebar bidang (*bandwidth*) dan menggeser frekuensi pusat ke nilai frekuensi yang lebih rendah. Sebaliknya pengurangan skala menurunkan durasi waktu, menambah lebar bidang dan menggeser frekuensi ke nilai frekuensi yang lebih tinggi (Burrus *et al*,1998).

Perapatan ($0 < a < 1$) dan peregangan ($a > 1$) akan menskala tanggapan frekuensi wavelet yang dibangkitkan, sehingga menghasilkan sejumlah wavelet yang mencakup rentang frekuensi yang diinginkan. Kumpulan wavelet ini dapat dianggap sebagai suatu bank tapis (*filter bank*) untuk analisis sinyal.

Keuntungan Transformasi Wavelet adalah bahwa jendelanya bervariasi. Untuk mengisolasi ketidakkontinuan sinyal, dapat digunakan fungsi basis yang sangat pendek. Pada saat yang sama, untuk analisis frekuensi secara terperinci, dapat digunakan fungsi basis yang sangat panjang.

Secara garis besar, Transformasi Wavelet dibedakan menjadi 2, yaitu Transformasi Wavelet Kontinu (*Continuous Wavelet Transform/CWT* atau *Integrated Wavelet Transform/IWT*) dan Transformasi Wavelet versi diskret. Versi diskret ada yang bersifat semi-diskret yang dikenal dengan runtun wavelet (*wavelet series*), dan ada yang diskret penuh yang dikenal dengan Transformasi Wavelet Diskret (*Discret Wavelet Transform/DWT*).

Pada Transformasi Wavelet Kontinu, waktu t serta parameter penskala a dan penggeser b berubah secara kontinu (dengan $a \neq 0$). Transformasi Wavelet kontinu fungsi $f(t)$ didefinisikan dengan :

$$\begin{aligned} TWK_f(a,b) &= \langle \psi_{a,b}(t), f(t) \rangle \\ &= \int_{-\infty}^{\infty} \psi_{a,b}(t) f(t) dt \\ &= \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) f(t) dt \end{aligned} \quad (2)$$

Transformasi Wavelet Kontinu ini mempunyai dua kelemahan, yaitu *redundancy* dan ketidakpraktisan (*impracticality*) (Mallat,1999). Masalah tersebut dapat diselesaikan dengan mendiskretkan parameter a dan b .

Pada transformasi yang bersifat semi diskret, dilakukan pendiskretan terhadap parameter a dan b, dengan $a = a_0^j$ dan $b = a_0^j k b_0$ dimana j dan k bilangan bulat, serta $a_0 > 1$ dan $b_0 > 0$. Pemilihan nilai a_0 b_0 bergantung pada wavelet ψ yang berkaitan dengan :

$$\psi_{j,k}(t) = \frac{1}{\sqrt{a_0^j}} \psi\left(\frac{t - kb_0 a_0^j}{a_0^j}\right) = \frac{1}{\sqrt{a_0^j}} \psi\left(\frac{t}{a_0^j} - kb_0\right)$$

Pendiskretan a dan b menghasilkan Runtun Wavelet/RW, yaitu :

$$RW(a_0^j, a_0^j k b_0) = \frac{1}{\sqrt{a_0^j}} \int_{-\infty}^{\infty} \psi\left(\frac{t}{a_0^j} - kb_0\right) f(t) dt$$

dengan membuat waktu t menjadi diskret maka diperoleh Transformasi Wavelet Diskret (TWD), yaitu :

$$TWD(a_0^j, a_0^j k b_0) = \frac{1}{\sqrt{a_0^j}} \sum_n \psi\left(\frac{n}{a_0^j} - kb_0\right) f(n)$$

Untuk $a_0 = 2$ dan $b_0 = 1$, waveletnya disebut *dyadic*, yaitu :

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t}{2^j} - k\right)$$

Transformasi Wavelet diskretnya menjadi

$$TWD_{dy}(2^j, 2^j k) = \frac{1}{\sqrt{2^j}} \sum_n \psi\left(\frac{n}{2^j} - k\right) f(n)$$

Dalam wavelet *dyadic* ini, skala j disebut dengan *oktaf* (Krisnan,1994).

Persamaan diatas merupakan hasil turunan dari persamaan Transformasi Wavelet Kontinu sehingga parameter skala kontinu a_0^j adalah kebalikan (*inversely*)

dari parameter skala *binary* 2^j dan hal inilah yang membuat 2^j muncul dalam persamaan diatas sebagai penyebut (*denominator*). Untuk itu Transformasi Wavelet Diskret yang digunakan adalah persamaan berikut(Gonzales,2002):

$$TWD_{dy}(2^j, 2^j k) = 2^{j/2} \sum_n \psi(2^j n - k) f(n)$$

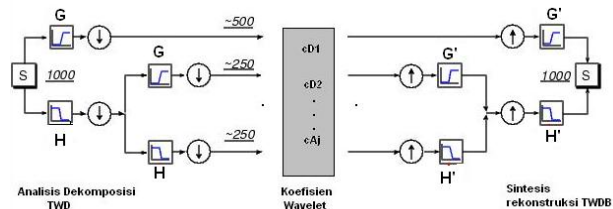
Sebuah teknik yang efisien untuk mengimplementasikan TWD adalah teknik analisis resolusi jamak (*multi resolution analysis*) yang dikembangkan Mallat tahun 1988. Analisis ini membawa kepada Transformasi Wavelet Cepat/TWC (Fast Wavelet Transform/FWT), dan diimplementasikan menggunakan *filter bank*.

Sinyal masukan S dilewatkan melalui 2 filter komplementer (*low-pass H* dan *high-pass G*), dan *downsampling* dengan membuang setiap data kedua, sehingga diperoleh koefisien pendekatan cA (komponen frekuensi rendah) dan koefisien detail cD (komponen

frekuensi tinggi). Proses ini dapat diiterasi dengan cara melanjutkan dekomposisi terhadap koefisien cA. Dengan demikian suatu sinyal dapat dipecah (didekomposisi) menjadi komponen-komponen dengan resolusi yang lebih rendah.

Proses sintesis sebagai kebalikan dari analisis bertujuan merekonstruksi sinyal masukan S, koefisien-koefisien cA dan cD dengan *upsampling* dan *filtering* (dengan filter H' dan G'). *Upsampling* merupakan proses penyisipan nilai nol antar dua data. Teknik rekonstruksi ini dapat dipeluas untuk komponen-komponen analisis *multi-resolusi* sampai pada tingkat tertentu.

Proses dekomposisi merupakan bagian analisis sinyal dengan Transformasi Wavelet Diskret (TWD), dan rekonstruksi yang merupakan bagian sintesis sinyal dengan Transformasi Wavelet Diskret Balik (TWDB) bertingkat sampai *oktaf* tertentu, secara lengkap terlihat pada **Gambar 5**.

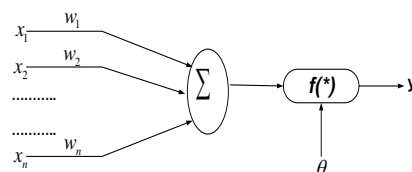


Gambar 5. Dekomposisi dan rekonstruksi *multistep*.

Jaringan Syaraf Tiruan (JST)

Dalam pengenalan pembicara *per* pembentukan model referensi pembicara dan pencocokan pola adalah dua tahapan yang sangat berkaitan. Pembentukan model referensi pembicara akan membentuk suatu model referensi yang akan digunakan untuk pencocokan pola. Salah satu teknik yang dapat digunakan dalam pencocokan pola adalah Jaringan Syaraf Tiruan (JST). Jaringan Syaraf Tiruan akan melakukan pembelajaran untuk membentuk suatu model referensi, kemudian JST yang telah melakukan pembelajaran tersebut dapat digunakan untuk pencocokan pola.

Jaringan Syaraf Tiruan didefinisikan sebagai sistem komputasi yang didasarkan pada pemodelan syaraf biologi (*neuron*) melalui pendekatan dari sifat-sifat komputasi biologis (*biological computation*). Jaringan Syaraf Tiruan bisa dibayangkan berupa jaringan dengan elemen pemroses sederhana yang saling terhubung. Seperti pada **Gambar 6**, elemen pemroses berinteraksi melalui sambungan *variable* yang disebut bobot, dan bila diatur secara tepat dapat menghasilkan sifat yang diinginkan (Fausett,1994).



Gambar 6. Sistem komputasi pemodelan *neuron*.

Dan pernyataan matematisnya

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right)$$

dengan x_i = sinyal masukan, $i = 1, 2, \dots, n$
 (n = banyaknya simpul masukan)
 w_i = bobot hubungan atau *synapsis*
 θ = *threshold* atau bias
 $f(*)$ = fungsi aktivasi
 y = sinyal keluaran dari *neuron*

Ide dasar JST adalah konsep belajar. Jaringan belajar melakukan generalisasi karakteristik tingkah laku obyek. Jika dilihat dari sudut pandang manusia, hal ini sama seperti bagaimana manusia belajar sesuatu. Manusia mengenal obyek dengan mengatur otak untuk menggolongkan atau melakukan generalisasi terhadap obyek tersebut.

Manusia menyimpan ilmu pengetahuannya ke dalam otak yang berisikan *synapsis*, *neuron*, dan komponen lainnya. Jaringan Syaraf Tiruan menyimpan ilmu pengetahuannya dalam nilai bobot sambungan (seperti *synapsis* dalam otak manusia) dan elemen-elemen (*neuron*) yang menghasilkan keluaran.

Untuk menyelesaikan permasalahan, JST memerlukan Algoritma untuk belajar, yaitu bagaimana konfigurasi JST dapat dilatih untuk mempelajari data histories yang ada. Dengan pelatihan ini, pengetahuan yang terdapat pada data bisa diketahui dan direpresentasikan dalam bobot sambungannya.

Jenis Algoritma belajar yang ada diantaranya (Jang *et al*, 1997) :

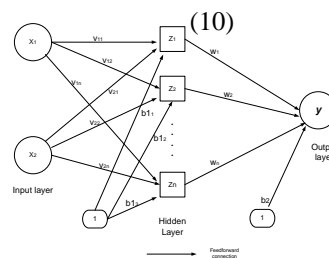
- a. *Supervised Learning*
 Algoritma ini diberikan target yang akan dicapai. Contohnya *Backpropagation Algorithm* (algoritma propagasi balik) dan *Radial Basis function*.
- b. *Unsupervised Learning*
 Pada Algoritma ini sama sekali tidak disediakan target, misalnya *Carpenter-Grossberg Adaptive Resonance Theory* (ART), *Learning Vector Quantization* (LVQ) dan *Competitive Learning Algorithm*.
- c. *Reinforcement Learning*
 Bentuk khusus dari supervised learning, contohnya *Genetic Algorithm* (GA).

Backpropagation (propagasi balik) yang merupakan salah satu model JST untuk pencocokan pola (*pattern matching*), menggunakan arsitektur *multi layer perceptron* (Gambar 7) dan pembelajaran propagasi balik. Walaupun JST propagasi balik membutuhkan waktu yang lama untuk pembelajaran tetapi bila pembelajaran telah selesai dilakukan, JST akan dapat mengenali suatu pola dengan cepat.

Beberapa karakteristik dari JST propagasi balik adalah sebagai berikut :

- ❖ *Jaringan Multi Layer*.
 JST propagasi balik mempunyai lapisan input, lapisan tersembunyi dan lapisan output dan setiap

neuron pada satu lapisan menerima input dari semua *neuron* pada lapisan sebelumnya.



Gambar 7. Arsitektur jaringan propagasi balik.

- ❖ *Fungsi Aktivasi*.
 Fungsi aktivasi akan menghitung input yang diterima oleh suatu *neuron*, kemudian *neuron* tersebut meneruskan hasil dari fungsi aktivasi ke *neuron* berikutnya, sehingga fungsi aktivasi berfungsi sebagai penentu kuat lemahnya sinyal yang dikeluarkan oleh suatu *neuron*. Beberapa fungsi aktivasi yang sering digunakan dalam JST propagasi balik adalah :

- Fungsi *sigmoid biner* yaitu fungsi biner yang memiliki rentang 0 s/d 1 dengan fungsi sebagai berikut :

$$f(x) = \frac{1}{1 + \exp(-x)}$$

- Fungsi *sigmoid bipolar*, yaitu fungsi yang memiliki rentang -1 s/d 1 dengan fungsi sebagai berikut :

$$f(x) = \frac{2}{1 + \exp(-x)} - 1$$

- ❖ Algoritma pembelajaran JST propagasi balik bersifat *iterative* dan didesain untuk meminimalkan *mean square error* (MSE) antara *output* yang dihasilkan dengan *output* yang diinginkan (target). Langkah-langkah Algoritma pembelajaran JST propagasi balik yang diformulasikan oleh Rumelhart, Hinton dan Rosenberg tahun 1986, secara singkat adalah sebagai berikut :

- ✚ Inisialisasi bobot, dapat dilakukan secara acak atau melalui metode Nguyen Widrow
- ✚ Perhitungan nilai aktivasi, tiap *neuron* menghitung nilai aktivasi dari input yang diterimanya. Pada lapisan input nilai aktivasi adalah fungsi identitas. Pada lapisan tersembunyi dan output nilai aktivasi dihitung melalui fungsi aktivasi
- ✚ Penyesuaian bobot, penyesuaian bobot dipengaruhi oleh besarnya nilai kesalahan (*error*) antara target *output* dan nilai *output* jaringan saat ini.
- ✚ Iterasi akan terus dilakukan sampai kriteria *error* tertentu dipenuhi.

Untuk mengimplementasikan algoritma diatas (pembelajaran), JST harus memiliki suatu set data pembelajaran. Data pembelajaran harus mencakup seluruh

jenis pola yang ingin dikenal agar nantinya dapat mengenali seluruh pola yang ada. Dalam kaitannya dengan sistem pengenalan pembicara, data pembelajaran harus mencakup seluruh pembicara yang ada. Dalam JST semakin banyak contoh suatu pola dalam pembelajaran maka JST akan semakin baik mengenal pola tersebut. Untuk itu akan lebih baik jika tiap pembicara mengucapkan lebih dari satu kali pengulangan untuk nantinya digunakan dalam pembelajaran JST.

JST akan menerima data input berupa vektor. Jika dimensi vektor terlalu besar maka JST akan bekerja lebih lambat. Dalam identifikasi pembicara setiap sinyal suara digital akan diproses terlebih dahulu dengan teknik ekstraksi ciri sehingga dimensi data akan tereduksi.

Dalam pembelajaran seluruh *set* data pembelajaran akan diproses sehingga JST akan membentuk suatu model referensi bagi seluruh pola-pola yang ada.

Pembuatan Keputusan

Pembuatan keputusan di dalam sistem pengenalan pembicara mencakup sistem identifikasi yang terkait erat dengan teknik pencocokan pola yang digunakan. Pembuatan keputusan dengan sistem identifikasi dapat dianalogikan dengan klasifikasi pola dengan tiap kelas merepresentasikan tiap pembicara. Pembuatan keputusan dapat dilakukan dengan metode nilai maksimum. Jika *neuron* output ke-*n* merupakan *neuron* dengan nilai maksimum maka data yang masuk dikenali sebagai pembicara ke-*n*.

HASIL DAN PEMBAHASAN

Pengujian yang telah dilakukan merupakan (1). Tahap praproses meliputi proses akuisisi data suara digital, proses frame blocking dan windowing, serta proses ekstraksi ciri dengan wavelet tipe orthogonal yaitu *Coiflet* dengan orde 4 pada level 10 dan 15. Pada proses akuisisi data suara digital dan proses frame blocking serta windowing, kombinasi suara yang digunakan dengan lamanya durasi perekaman, memberikan hasil untuk tiap data suara berupa vektor, dengan dimensi 66.150 (22.050 x 3).

Jumlah pembicara yang diambil sebagai data training adalah 50 (lima puluh) data suara yang dipilih dengan index ganjil dari 100 (seratus) data hasil perekaman, kemudian sisanya dengan index genap digunakan untuk data test identifikasi. Hal ini dilakukan agar menghasilkan pengenalan yang lebih akurat, mengingat sinyal suara sangat terpengaruh (*sensitive*) terhadap waktu.

Proses Ekstraksi ciri dari *Coiflet* orde 4 (db4) pada level 10 memberikan hasil tiap satu data *file* (satu file *.wav) pembicara (dari 10 data *file* setiap pembicara) menghasilkan koefisien sebanyak 150 elemen dengan *elapsed time* praprosesnya rata-rata sebesar 1,151 detik, dan pada saat 50 data suara yang digunakan untuk training diproses, membutuhkan waktu 80,465 detik. Dari 50 data suara tersebut akan membentuk suatu matriks dengan dimensi 134 x 50, yang digunakan sebagai input untuk proses pembentukan referensi pembicara dan pencocokan pola. Sedangkan pada level 15, satu *file* pembicara (dari 10 *file* setiap pembicara) menghasilkan koefisien sebanyak 26 elemen dengan *elapsed time*

praprosesnya sebesar 1,141 detik, sedangkan untuk 50 data yang digunakan untuk training menghabiskan waktu sebesar 88,307 detik.

Pengujian selanjutnya adalah (2). Tahap referensi pembicara dan pencocokan pola dengan menggunakan laju pembelajaran 0.01 dan 0.3 serta toleransi galat sebesar 0.00001. Kemudian untuk menentukan banyaknya neuron tersembunyi (*hidden neuron*) menggunakan metode *trial and error* dari 10 sampai dengan 100 neuron. Sehingga diperoleh hasil, (a). Pada pengujian yang dilakukan pada tipe wavelet *coif4* level 10, dengan toleransi galat 0.00001, laju pembelajaran 0.01 dan mencapai pengenalan tertinggi pada jumlah *neuron* tersembunyi 20, dengan waktu 7.57 detik. Generalisasi yang dihasilkan terhadap data testing sebesar 84% (42 pembicara dari 50 data pembicara dikenali). Sedangkan (b) Pada pengujian *Coiflet* orde 4 (*coif4*) level 15 memberikan hasil, laju pembelajaran 0.01 dan *neuron* tersembunyi 100 memberikan tingkat pengenalan yang paling baik, namun proses pembelajarannya berjalan sangat lambat dan kinerja tujuannya (10^{-5}) secara keseluruhan belum tercapai. Generalisasi terbaik yang dicapai yaitu 78%, menurun dibandingkan dengan *coif4* level 10 yang memiliki koefisien lebih banyak. Hal ini menunjukkan bahwa untuk proses ekstraksi ciri menggunakan metode Transformasi Wavelet, semakin tinggi tingkat (level) dekomposisi, yang berimplikasi terhadap jumlah koefisiennya (koefisien yang dihasilkan semakin tereduksi) maka proses pembelajaran JST berjalan sangat lambat dan menurunkan tingkat pengenalan jaringannya (dapat ditunjukkan pada hasil generalisasi).

KESIMPULAN DAN SARAN

Transformasi Wavelet Diskret dapat digunakan dalam proses identifikasi pembicara pada bagian pemrosesan awal (praproses) sinyal untuk mendapatkan informasi (ciri) sinyal tersebut. Sistem identifikasi pada tahap praproses *coif4* level 10 menghasilkan tingkat generalisasi tertinggi sebesar 84%. Pada transformasi wavelet, tingkat (level) dekomposisi mempengaruhi tingkat pengenalan jaringan yaitu makin tinggi tingkat dekomposisi maka tingkat pengenalan jaringan semakin menurun. Jaringan syaraf tiruan propagasi balik baik digunakan untuk pembentukan referensi pembicara dan pencocokan pola.

Penelitian ini masih dapat dikembangkan lebih jauh dan lebih dalam lagi yang nantinya diharapkan dapat terbentuk suatu sistem yang lebih baik.

Saran-saran untuk penelitian lebih lanjut antara lain :

1. Perlu ada pengkajian lebih lanjut mengenai Transformasi Wavelet, untuk mendapatkan tingkat pengenalan yang lebih tinggi.
2. Melakukan penambahan jumlah pembicara untuk melihat kinerja sistem dengan jumlah data yang besar.
3. Penggunaan JST yang bersifat *incremental learning*, sehingga JST dapat mengenali pola baru dengan lebih baik.

DAFTAR PUSTAKA

- Burrus, C.S. Gopinath R.A., Guo, H. 1998, *Introduction to Wavelets and Wavelet Transforms A Primer*, International Edition, Prentice-Hall International, Inc.
- Campbell, J.P.,JR. 1997. *Speaker Recognition: A Tutorial*. Proc. IEEE, vol.85, no 9,pp1437-1462,1997
- Fausett, L. 1994. *Fundamentals of Neural Network*. Prentice Hall, Englewood Cliffs,NJ.
- Fu, L. 1994. *Neural Network in Computer Intelligence*. McGraw-Hill, Singapore.
- Gonzalez R.C., Wood, R.E. 2002, *Digital Image Processing*, Second Edition, Prentice Hall, Inc
- Ho, C.E. 1998. *Speaker Recognition System*. Project Report. California Institut of Technology.
- Jang,J.S.R, Sun, C.T, Mizutani, E. 1997, *Neuro Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligence*, International Edition, Prentice-Hall International Inc.
- Krishnan, M., Neophytou C.P., Prescott G.1994, *Wavelet Transform Speech Recognition Using Vector Quantization, Dynamic Time Warping and Artificial Neural Networks*, Center of excellence in computer aided systems engineering and Telecommunication & Information Sciences Laboratory 2291 Irving Hill Drive, Lawrence, KS 66045.
- Mallat, Stephane, 1999, *A Wavelet Tour Of Signal Processing*, Second Edition, Academic Press 24-28 Oval Road, London NW1 7DX UK, <http://www.hbuk.co.uk/ap/>
- Mudry, A.H.1997. *Speaker Identification using Wavelet Transform*, Tesis Master of Engineering, Ottawa-Carleton Institut of Electrical Engineering, Ontario
- Proakis J.G., Manolakis.D.G. 1997, *Pemrosesan Sinyal Digital*, edisi Bahasa Indonesia jilid 1, Prenhallindo, Jakarta
- Rabiner, L.R., Juang B.H., 1993, *Fundamentals of Speech Recognition*, Prentice Hall, New Jersey, 1993, ISBN 0-13-015157-2.
- Xafopoulos, A. 2001. *Speaker Verification (an overview)*. TUT-TICSP presentation. TICSP (Tampere International Center for Signal Processing), TUT (Tampere Univ. of Technology), Tampere Finland.
- Johnson A.J., 1962, *On the Vibration an Amplitudes of Ship Hulls*, Transaction of IESS.
- Lewis F. M, 1929, *The Inertia of Water Surrounding of Vibration Ship*, Transaction of RINA., Volume 37, London.
- Schlick O., (1894 – 1991), *Summary of Article on Ship Vibration*, in Transaction of INA., London.
- Taylor J.L., 1930, *Some Measurement of Ship Vibration*, Transaction of NECL., Newcastle Upon Tyne.
- Todd F. M., 1961, *Ship Hull Vibration, A Comparises of Measured and Calculated Frequencies*, Edward Arnold (Publishers) Ltd., First Edition.
- Townsin N. L., 1961, *The Calculation of the Higher Modes Frequencies in Ship Vibration*, The Shipbuilder and Marine Engineer Builder, London.
- Vorus W. 1988, Chapter Entitled “*Vibration*”, The Principles of Naval Architecture, The Society of NAME., New York.