

## PERLUASAN METODE MFCC 1D KE 2D SEBAGAI ESKTRAKSI CIRI PADA SISTEM IDENTIFIKASI PEMBICARA MENGUNAKAN *HIDDEN MARKOV MODEL* (HMM)

Agus Buono<sup>1\*)</sup>, Wisnu Jatmiko<sup>2</sup>, dan Benyamin Kusumoputro<sup>2</sup>

1. Departemen Ilmu Komputer, FMIPA, IPB, Kampus IPB Darmaga, Bogor 16680, Indonesia

2. Laboratorium Kecerdasan Komputasional, Fakultas Ilmu Komputer, Universitas Indonesia, Depok 16424, Indonesia

<sup>\*)</sup>E-mail: pudesha@yahoo.co.id; nynykusumo@yahoo.com

---

### Abstrak

Pada paper ini, sebuah metode *Mel-Frequency Cepstrum Coefficients* (MFCC) sebagai pengekstraksi ciri telah dikembangkan sebagai perluasan dari MFCC pada ruang satu dimensi (1D-MFCC) ke ruang dua dimensi (2D-MFCC). Pada metode 1D-MFCC, komponen masukan sistem adalah nilai spectrum sinyal yang diperoleh melalui transformasi Fourier, maka pada metode 2D-MFCC sebagai komponen masukan sistem adalah data bispectrum sinyal. Oleh karena itu, perubahan yang dilakukan adalah dimensi filter serta transformasi kosinus pada tahap akhir metode dari ruang satu dimensi menjadi ruang dua dimensi. Efektifitas metode 2D-MFCC yang diusulkan ini kemudian diuji pada sistem untuk identifikasi pembicara menggunakan HMM sebagai pengenalan pola. Percobaan dilakukan pada sistem untuk mengenali 10 pembicara dengan data latihan masing-masing pembicara sebanyak 20, 40 dan 60 data suara dari 80 suara yang tersedia untuk setiap pembicara. Hasil dengan data uji menunjukkan bahwa sistem memiliki akurasi yang tinggi, yaitu diatas 99%, 92% dan 88% dengan masing-masing untuk model dengan data latihan sebanyak 60, 40 dan 20.

### Abstract

The Extention of MFCC Technique from 1D to 2D as Feature Extractor for Speaker Identification System Using HMM. In this paper, we introduce an extension of Mel-Frequency Cepstrum Coefficients (1D-MFCC) methodology to bispectrum data, referred to as 2D-MFCC, for feature extraction. 2D-MFCC is based on 2D bispectrum data rather than 1D spectrum vector yielded by Fourier transform, so the filter in 1D-MFCC must be extend to 2D filter and using 2D cosine transform to get the mel-cepstrum coefficients from the filtered bispectrum values. Based on 2D-MFCC, we develop a speaker recognition system with Hidden Markov Model (HMM) as classifier. The experimental results show that the recognition rate is around 88%, 92% and 99% for 20, 40 and 60 data training, respectively.

*Keywords: Mel-Frequency Cepstrum Coefficients (MFCC), Bispectrum, Hidden Markov Model (HMM)*

---

### 1. Pendahuluan

Mel-Frequency Cepstrum Coefficients (MFCC) merupakan teknik ekstraksi ciri yang telah luas dipakai pada pemrosesan sinyal suara, terutama pada pengenalan pembicara. Penggunaan teknik ini pada sistem pemrosesan sinyal memberikan pengenalan yang lebih baik dibandingkan dengan metode lain yang sudah ada [1].

Sejak diperkenalkan oleh Davis and Mermelstein pada tahun 1980 [1], maka teknik MFCC ini telah mengalami pengembangan dan variasi. Bagian utama yang mengalami modifikasi adalah pada pengembangan

filter, seperti jumlah, bentuk, cara membuat sekatan, serta lebar filter yang dibuat. Sedangkan dari sisi masukan tetap sama, yaitu spektrumenergi dari input yang diperoleh melalui transformasi Fourier. Dari studi yang sudah dilakukan [2], terlihat bahwa spektrum ini sensitif terhadap gangguan noise. Hal ini berakibat sistem yang dikembangkan akan mempunyai akurasi yang rendah untuk sinyal yang terkontaminasi oleh noise. Dengan noise 20 dB, akurasi sistem turun menjadi sekitar 40% dari semula 99% untuk data tanpa penambahan noise [2]. Oleh karena itu perlu dicari besaran lain yang akan menggantikan spektrum energi yang menjadi komponen masukan pada metode MFCC ini.



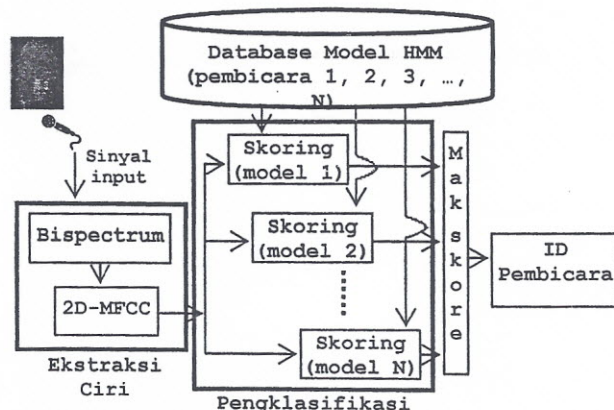
Pada [3] disebutkan bahwa nilai spektrum orde tinggi *Higher Order Statistics* (HOS) untuk orde 3 (bispektrum) dari sinyal yang berdistribusi Gaussian adalah nol. Fakta ini diharapkan mampu menekan pengaruh Gaussian noise pada sinyal, yang pada akhirnya sistem yang dikembangkan akan bersifat robust terhadap noise. Secara empiris, hal ini telah dibuktikan pada [4-6]. Oleh karena bispektrum adalah berdimensi dua, maka teknik MFCC konvensional perlu dikembangkan sehingga dapat memproses untuk data dua dimensi. Bertolak dari hal tersebut, maka pada paper ini, disajikan suatu perluasan metode MFCC dari satu dimensi (1D-MFCC) ke dua dimensi (2D-MFCC).

Selanjutnya, paper ini disajikan dengan susunan sebagai berikut: Bagian 2 mengenai prinsip sistem identifikasi pembicara. Metode MFCC untuk ekstraksi ciri pada bagian 3. Sedangkan perluasannya pada bagian 4. Bagian 5 menyajikan rancangan percobaan serta hasilnya. Sistem yang dikembangkan untuk melihat efektifitas teknik yang diusulkan ini terdiri dari 10 pembicara. Akhirnya, kesimpulan serta saran untuk penelitian selanjutnya disajikan pada bagian 6.

## 2. Metode Penelitian

**Prinsip Sistem Identifikasi Pembicara.** Identifikasi pembicara merupakan proses untuk menentukan pembicara berdasar input suara yang diberikan [7]. Secara umum, sistem identifikasi pembicara terdiri dari dua subsistem, yaitu subsistem ekstraksi ciri dan subsistem pencocokan pola, seperti disajikan pada Gambar 1.

Subsistem ekstraksi ciri melakukan proses transformasi sinyal input ke dalam satu set vektor ciri sebagai representasi dari sinyal suara suatu pembicara untuk proses selanjutnya. Subsistem pencocokan pola merupakan bagian untuk melakukan identifikasi suatu pembicara yang belum diketahui dengan cara membandingkan sinyal suaranya yang telah diekstrak ke



Gambar 1. Blok Diagram Sistem Identifikasi Pembicara dengan HMM Sebagai Pengklasifikasi

dalam vektor ciri dengan set vektor ciri dari pembicara yang telah diketahui dan tersimpan dalam sistem [7].

Dari aspek pengembangan sistem, ada dua fase pada sistem identifikasi pembicara. Fase pertama adalah tahap pelatihan. Pada fase ini sistem melakukan pelatihan untuk menentukan parameter model untuk setiap pembicara berdasar data suara pembicara tersebut. Pada penelitian, frase yang digunakan adalah "pudsha". Pada penelitian ini setiap pembicara dimodelkan dengan menggunakan *Hidden Markov Model* (HMM). Dari sampel data dengan frase "pudsha" ini, model setiap pembicara dilatih dengan menggunakan algoritma Baum Welch seperti yang disajikan pada [8].

Fase kedua adalah tahapan pengujian, yaitu sinyal input yang diberikan kepada sistem dicocokkan dengan model setiap pembicara yang ada pada sistem. Keputusan untuk menentukan pembicara didasarkan pada skor tertinggi untuk setiap model. Untuk penghitungan skor ini digunakan algoritma *Forward* [8].

**Mel-Frequency Cepstrum Coefficients.** Sinyal suara adalah sinyal yang berubah terhadap waktu secara perlahan (*quasi-stationary*). Untuk rentang waktu yang pendek (antara 5 and 100 milliseconds), karakteristik sinyal ini bisa dianggap stasioner. Namun demikian untuk periode yang cukup panjang (1/5 seconds atau lebih) karakteristik sinyal berubah yang mencerminkan perbedaan bunyi yang diucapkan. Oleh karena itu, *short-time spectral analysis* adalah yang biasa dilakukan untuk mengkarakterisasi sinyal suara [9].

MFCC adalah teknik ekstraksi ciri yang populer dan paling banyak digunakan. Block diagram teknik MFCC ini disajikan pada Gambar 2 [9].

**Frame blocking:** sinyal suara dibaca per blok (atau frame, terdiri S sample), yang digeser dari awal hingga akhir. Antara dua frame yang *adjacent* terdapat overlap S-L sample (dengan kata lain L adalah lebar pergeseran frame, dengan  $L < S$ ).

**Windowing:** proses windowing dilakukan pada setiap frame dengan tujuan untuk meminimumkan diskontinuitas antar dua frame yang *adjacent*, khususnya pada bagian awal dan akhir.

**FFT:** Pada tahap ini setiap frame yang terdiri dari N samples dikonversi dari domain waktu ke domain frekuensi. Output dari proses ini disebut dengan nama spektrum atau periodogram.

**Mel-Frequency wrapping:** tahap ini merupakan proses pengfilteran dari spektrum setiap frame yang diperoleh dari tahapan sebelumnya, menggunakan sejumlah M filter segitiga dengan tinggi satu. Filter ini dibuat

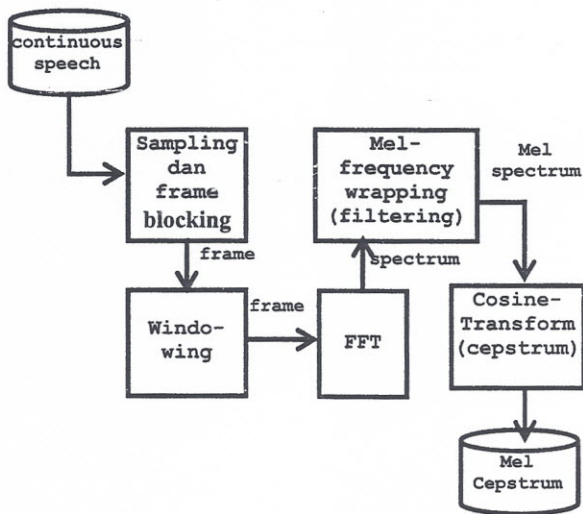


dengan mengikuti persepsi telinga manusia dalam menerima suara. Persepsi ini dinyatakan dalam skala 'mel' (berasal dari Melody) yang mempunyai hubungan tidak linear dengan frekuensi suara, [9]. Dalam hal ini skala mel-frequency adalah linear untuk frekuensi kurang dari 1000 Hz dan logaritmik untuk frekuensi di atas 1000 Hz. Satu relasi antara frekuensi bunyi (dalam Hz) dengan skala mel adalah, [7,9]:

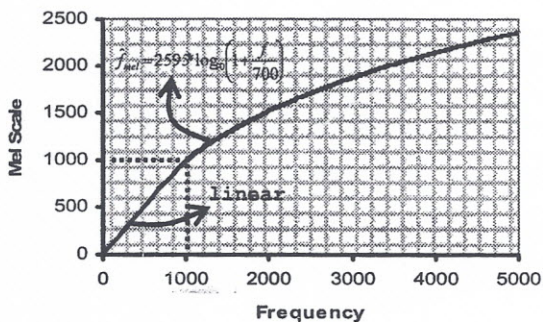
$$\hat{f}_{mel} = 2595 * \log_{10} \left( 1 + \frac{f}{700} \right) \quad (1)$$

Grafik relasi di atas disajikan pada Gambar 3.

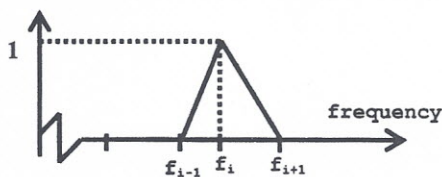
Alur pembuatan M filter tersebut sesuai dengan algoritma 1. Sedangkan Gambar 4 menyajikan contoh sebuah filter segitiga ke-i.



Gambar 2. Block Diagram Teknik MFCC



Gambar 3. Hubungan antara Frekuensi dengan Skala Mel



Gambar 4. Suatu Filter Segitiga ke i dengan Tinggi 1

**Algoritma 1: Mengkonstruksi M filter**

- a. Pilih jumlah filter yang akan dibuat (M)
- b. Pilih frekuensi terbesar ( $f_{high}$ ). Dari nilai ini, maka nilai tertinggi dari  $\hat{f}_{mel}$  adalah

$$\hat{f}_{mel}^{high} = 2595 * \log_{10} \left( 1 + \frac{f_{high}}{700} \right)$$

- c. Pusat filter ke i adalah  $f_i$ :
  - c.1.  $f_i = \frac{1000}{0.5 * M} * i$  untuk  $i=1, 2, 3, \dots, M/2$
  - c.2. untuk  $i=M/2, M/2+1, \dots, M$ , maka  $f_i$  dihitung dengan prosedur berikut:
    - 1. skala mel disekat dengan lebar yang sama, yaitu sebesar  $\Delta$ , dengan:

$$\Delta = \frac{\hat{f}_{mel}^{high} - 1000}{0.5 * M}$$

Dengan persamaan (1), maka nilai  $\Delta$  dapat dirumsukan sebagai :

$$\Delta = \frac{5190}{M} \log \left( \frac{700 + f_{high}}{1700} \right)$$

- 2. nilai mel untuk pusat filter ke i adalah:
 
$$a = 1000 + (i - 0.5 * M) * \Delta$$
- 3. pusat dari filter ke i adalah:
 
$$f_i = 700 * (10^{a/2595} - 1)$$

Nilai koefisien spektrum mel (*mel spectrum coefficients*) adalah jumlah dari hasil pengfilteran sesuai rumus 2, [7]:

$$X_i = \log \left( \sum_{j=0}^{N-1} abs(X(j)) * H_i(f) \right) \quad (2)$$

Dengan  $i = 1, 2, 3, \dots, M$ ; M adalah banyaknya filter; N adalah banyaknya koefisien FFT;  $abs(X(j))$  adalah magnitudo koefisien ke j dari periodogram yang dihasilkan oleh transformasi FFT; dan  $H_i(f)$  adalah nilai filter ke i pada titik frekuensi f.

*Cepstrum*: Pada tahap ini dilakukan konversi dari koefisien spektrum mel kembali ke domain waktu menggunakan transformasi kosinus berikut :

$$C_j = \sum_{i=1}^M X_i * \cos \left( \frac{j * (i - 0.5) * \pi}{20} \right) \quad (3)$$

dengan  $j = 1, 2, 3, \dots, K$ ; K adalah banyaknya koefisien MFCC yang diinginkan; M adalah banyaknya filter segitiga;  $X_i$  adalah koefisien spektrum mel yang diperoleh dengan Pers. (2). Dalam hal ini  $C_j$  disebut sebagai koefisien ke j dari *mel frequency cepstrum coefficients* (MFCC).

**Perluasan MFCC dimensi 1 ke 2.** Pada banyak aplikasi pemrosesan sinyal, ekstraksi ciri dilakukan pada setiap frame. Untuk sistem dengan 1D-MFCC sebagai pengekstraksi ciri, komponen yang diproses dari setiap frame adalah nilai spektrum atau periodogram. Periodogram ini diperoleh dari transformasi Fourier dan berupa vektor berdimensi 1 dengan N elemen. Pada 2D-MFCC, periodogram tersebut digantikan oleh bispektrum.



Bispektrum ini berupa array dimensi 2 dengan ukuran  $N \times N$ , dan setiap unsurnya dinotasikan dengan  $B(f1, f2)$ . Dalam hal ini  $B(f1, f2)$  adalah nilai bispektrum pada frekuensi  $(f1, f2)$ , dengan  $f1$  dan  $f2$  adalah  $1, 2, 3, \dots, N$ . Pada bagian ini akan disajikan sekilas tentang bispektrum, dilanjutkan dengan mengformulasikan metode yang diusulkan untuk memperluas metode 1D-MFCC ke 2D-MFCC.

**Bispektrum.** Jika  $\{X(k)\}$ ,  $k = 0, \pm 1, \pm 2, \dots$ , adalah proses stokastik yang bernilai real, maka *cumulant* order 3 adalah  $c_3^x(\tau_1, \tau_2)$ , yang dirumuskan sebagai, [8]:

$$c_3^x(\tau_1, \tau_2) = \sum_{p=1}^3 \sum_{m=1}^p (-1)^{p-1} (p-1)! E \left( \prod_{k \in s_1} X_k \right) E \left( \prod_{k \in s_2} X_k \right) E \left( \prod_{k \in s_3} X_k \right) \quad (4)$$

R adalah banyaknya cara menyekat set  $\{X_k, X_{k+1}, X_{k+2}\}$  menjadi p sekatan, dengan  $p = 1, 2, 3$ . Sebagai ilustrasi, untuk  $p = 2$ , maka diperoleh 3 kemungkinan sekatan ( $R = 3$ ), yaitu:  $s_1 = \{X_k, X_{k+1}\}$ ,  $s_2 = \{X_{k+2}\}$ ;  $s_1 = \{X_k\}$ ,  $s_2 = \{X_{k+1}, X_{k+2}\}$ ; dan  $s_1 = \{X_{k+1}\}$ ,  $s_2 = \{X_k, X_{k+2}\}$ . Bispektrum, yang disebut juga sebagai spektrum *cumulant*, adalah transformasi Fourier dari barisan *cumulant* tersebut, dan diformulasikan sebagai [8]:

$$C_3^x(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} c_3^x(\tau_1, \tau_2) \exp\{-j(\omega_1 \tau_1 + \omega_2 \tau_2)\} \quad (5)$$

Untuk proses stasioner, *cumulant* order 3 dapat diformulasikan sebagai:

$$c_3^x(\tau_1, \tau_2) = E\{x(t)x(t+\tau_1)x(t+\tau_2)\} \quad (6)$$

Paramater  $\tau_1$  dan  $\tau_2$  pada Pers. (4-6) di atas adalah lag yang secara teoritis bernilai bilangan real. Pada prakteknya, nilai bispektrum ini diduga dari sejumlah samples data. Secara umum ada dua pendekatan dalam menduga bispektrum, yaitu pendekatan parametrik dan pendekatan konvensional. Pendekatan konvensional dikelompokkan menjadi tiga, yaitu teknik tidak langsung (*indirect technique*), teknik langsung (*direct technique*), dan modulasi kompleks (*complex demodulates*). Pada penelitian ini digunakan metode konvensional dengan teknik tidak langsung untuk menduga nilai bispektrum. Hal ini dikarenakan teknik ini lebih sederhana dibanding lainnya. Algoritma secara lengkap dapat dilihat pada [8].

**Formulasi Perluasan 1DMFCC ke 2DMFCC.** Pada 2D-MFCC, komponen yang masuk ke proses filtering adalah nilai bispektrum. Oleh karena bispektrum ini berdimensi 2, maka filter pada 1D-MFCC dikembangkan menjadi berdimensi 2. Berikut adalah algoritma untuk membentuk filter dimensi dua:

**Algoritma 2:** Pembentukan filter dimensi dua

- a. Menggunakan algoritma 1, tentukan pusat filter ke  $i$  untuk sumbu F1 (yaitu  $f1_i$ ) dan juga untuk sumbu F2 (yaitu  $f2_i$ ).

- b. Filter dimensi dua berupa limas dengan tinggi satu dan proyeksi titik puncak ke bidang alas pada titik  $(f1_i, f2_j)$ . Sudut dari bidang alasnya adalah pada titik  $(f1_{i-1}, f2_{j-1})$ ,  $(f1_{i+1}, f2_{j-1})$ ,  $(f1_{i-1}, f2_{j+1})$ , dan  $(f1_{i+1}, f2_{j+1})$ .

Dengan filter dimensi dua ini, nilai koefisien *mel-spektrum* (*mel-spectrum coefficients*) pada Pers. (2) akan berubah menjadi:

$$X(i, j) = \log \left[ \sum_{f1=1}^{128} \sum_{f2=1}^{128} B(f1_m, f2_n) * H_{i,j}(f1_m, f2_n) \right] \quad (7)$$

Nilai ini disebut sebagai koefisien *mel-bispektrum* (*mel-bispectrum coefficients*). Pada Pers. (7) di atas,  $X(i, j)$  adalah nilai koefisien *mel-bispektrum* untuk filter  $H_{i,j}$ , dengan  $i, j = 1, 2, 3, \dots, M$ , dan  $M$  adalah banyaknya sekatan untuk filter pada sumbu F1 atau F2.  $B(f1_m, f2_n)$  adalah nilai bispektrum pada titik  $(f1_m, f2_n)$ . Sementara itu  $H_{i,j}(f1_m, f2_n)$  adalah tinggi filter  $H_{i,j}$  pada titik  $(f1_m, f2_n)$ , dan diformulasikan dengan Algoritma 3. Sedangkan, Gambar 5 menyajikan ilustrasi alir algoritma tersebut.

**Algoritma 3:** Penghitungan  $H_{i,j}$  di titik  $(f1_m, f2_n)$

- a. Bidang alas filter  $H_{i,j}$  disekat menjadi empat yang merupakan proyeksi dari setiap sisi filter limas ke bidang alas.
- b. Tentukan batas-batas daerah pada tahap a.
- c. Tentukan daerah mana titik  $(f1_m, f2_n)$  berada dengan aturan berikut:

Jika:

- a.  $f2_n > f2_{j-1}$ , and
- b.  $f2_n < \left( \frac{f2_{j-1} - f2_j}{f1_{i-1} - f1_i} \right) (f1_m - f1_i) + f2_j$
- c.  $f2_n < \left( \frac{f2_{j-1} - f2_j}{f1_{i+1} - f1_i} \right) (f1_m - f1_i) + f2_j$

Maka  $(f1_m, f2_n) \in A$

Jika:

1.  $f2_n < f2_{j+1}$ , and
2.  $f2_n > \left( \frac{f2_{j+1} - f2_j}{f1_{i+1} - f1_i} \right) (f1_m - f1_i) + f2_j$
3.  $f2_n > \left( \frac{f2_{j+1} - f2_j}{f1_{i-1} - f1_i} \right) (f1_m - f1_i) + f2_j$

Maka  $(f1_m, f2_n) \in B$

Jika:

1.  $f1_m > f1_{i-1}$ , and
2.  $f2_n > \left( \frac{f2_{j-1} - f2_j}{f1_{i-1} - f1_i} \right) (f1_m - f1_i) + f2_j$
3.  $f2_n < \left( \frac{f2_{j+1} - f2_j}{f1_{i-1} - f1_i} \right) (f1_m - f1_i) + f2_j$

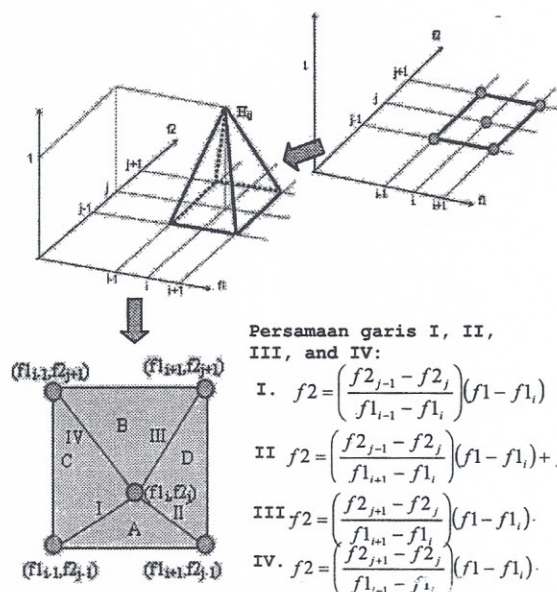
Maka  $(f1_m, f2_n) \in C$

Jika:

1.  $f1_m < f1_{i+1}$ , and
2.  $f2_n > \left( \frac{f2_{j-1} - f2_j}{f1_{i+1} - f1_i} \right) (f1_m - f1_i) + f2_j$
3.  $f2_n < \left( \frac{f2_{j+1} - f2_j}{f1_{i+1} - f1_i} \right) (f1_m - f1_i) + f2_j$

Maka  $(f1_m, f2_n) \in D$





Gambar 5. Konstruksi Filter pada Ruang Dimensi 2  $H_{ij}$  dengan Pusat Titik  $(f_{1i}, f_{2j})$

d. Hitung nilai  $H_{i,j}(f_{1m}, f_{2n})$  sesuai dengan daerahnya, dengan menggunakan aturan berikut:

Jika  $(f_{1m}, f_{2n}) \in A$  maka:  $H_{i,j}(f_{1m}, f_{2n}) = \frac{f_{2n} - f_{2j-1}}{f_{2i} - f_{2i-1}}$

Jika  $(f_{1m}, f_{2n}) \in B$  maka:  $H_{i,j}(f_{1m}, f_{2n}) = \frac{f_{2j+1} - f_{2n}}{f_{2i+1} - f_{2i}}$

Jika  $(f_{1m}, f_{2n}) \in C$  maka:  $H_{i,j}(f_{1m}, f_{2n}) = \frac{f_{1m} - f_{1i-1}}{f_{1i} - f_{1i-1}}$

Jika  $(f_{1m}, f_{2n}) \in D$  maka:  $H_{i,j}(f_{1m}, f_{2n}) = \frac{f_{1i+1} - f_{1m}}{f_{1i+1} - f_{1i}}$

Pada Gambar 5 terlihat bahwa bidang filter terbagi menjadi 4 daerah, yang masing-masing sebagai proyeksi sisi filter ke bidang alas, yang dipisahkan empat garis, yaitu garis I, II, III, dan IV.

Tahap akhir dari 2D-MFCC ini adalah mengkonversi koefisien *mel-bispectrum* yang berdimensi  $M \times M$  dari domain frekuensi menjadi dalam domain waktu. Kalau pada 1D-MFCC, teknik yang digunakan adalah transformasi kosinus orde satu, dengan Pers. (3), maka pada 2D-MFCC menggunakan Pers. (8).

$$Y(k) = \sum_{i=1}^M \sum_{j=1}^M X(i, j) * \cos\left(\frac{k(i-0.5)\pi}{M}\right) * \cos\left(\frac{k(j-0.5)\pi}{M}\right) \quad (8)$$

$Y(k)$  adalah koefisien ke  $k$  dari 2D-MFCC, dengan  $k = 1, 2, 3, \dots, K$ ;  $K$  adalah banyaknya koefisien, dan  $X(i, j)$  adalah koefisien *mel-bispectrum* dari Pers. (6).

### 3. Hasil dan Pembahasan

Metode yang diusulkan diterapkan pada sistem identifikasi pembicara yang melibatkan 10 pembicara.

Setiap pembicara diminta mengucapkan frase “pudsha” sebanyak 80 kali. Setiap sinyal suara yang dihasilkan disampling dengan durasi 1,28 detik pada sampling rate 11 kHz. Oleh karena itu, secara keseluruhan diperoleh  $10 \times 80 = 800$  data suara dari 10 pembicara. Dari 80 data untuk setiap pembicara dipisahkan menjadi 2 set, yaitu satu set sebagai data training dan sisanya sebagai data uji. Dalam hal ini dilakukan 3 jenis pembagian dengan rasio data training:data uji sebagai 60:20, 40:40, dan 20:60.

Berikutnya data suara pada setiap set dibaca dari frame demi frame dengan lebar frame 512 sample dan overlap antar frame 256 sample. Setiap frame yang dihasilkan dihitung nilai bispektrum untuk frekuensi  $128 \times 128$ . Dari sini dihitung nilai cirinya dengan menggunakan 2D-MFCC untuk mendapatkan 13 koefisien MFCC dari data bispektrum setiap frame tersebut.

Untuk mengetahui seberapa besar komponen koefisien MFCC dalam menyumbang terhadap kinerja sistem, dilakukan tiga skenario penggunaan koefisien MFCC, yaitu skenario 1 (menggunakan semua koefisien), skenario 2 (menghapus koefisien ke 1), skenario 3 (menghapus koefisien ke 1 dan ke 2). Setiap percobaan dibuat model HMM sebagai pengenalan pola dengan jumlah hidden state mulai dari 3 s/d 7. Oleh karena itu, secara keseluruhan jumlah percobaan yang dilakukan adalah sebanyak  $3 \times 3 \times 5 = 45$  (3 jenis rasio data training dengan data uji, 3 jenis skenario, dan 5 jenis jumlah *hidden state*). Untuk melihat perbandingan dengan teknik sebelumnya, maka juga dilakukan pengenalan suara menggunakan 1D-MFCC sebagai ekstraksi ciri.

Gambar 6 menyajikan perbandingan hasil pengenalan terhadap 10 pembicara antara dua metode, yaitu 1D-MFCC dengan 2D-MFCC untuk data suara tanpa penambahan noise. Terlihat bahwa kedua metode memberikan akurasi yang tinggi ( $> 99\%$ ) untuk data suara tanpa penambahan noise. Juga terlihat bahwa secara relatif teknik yang diusulkan mampu mendeteksi dengan akurasi lebih tinggi dibanding teknik sebelumnya, yaitu 1D-MFCC untuk semua jumlah hidden state yang dipilih. Gambar tersebut juga menunjukkan bahwa pemilihan jumlah *hidden state* tidak terlalu memberikan perbedaan berarti dalam hal tingkat akurasi.

Hal yang menarik adalah bahwa metode yang diusulkan mampu memberikan hasil akurasi mencapai 100% untuk jumlah hidden state sebanyak 6, yang dicapai dengan proporsi jumlah data training 75%.

Namun demikian, jika data suara diberi noise sebesar 20 dB, ternyata akurasi sistem turun sangat drastis menjadi sekitar 40% untuk kedua metode, seperti ditunjukkan pada Gambar 7. Untuk memperbaiki akurasi, perhatikan

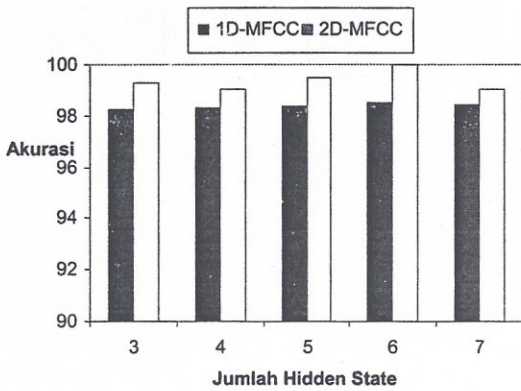


Gambar 8 yang menyajikan sensitifitas komponen MFCC terhadap noise. Dari gambar tersebut terlihat bahwa Komponen koefisien dari vektor ciri yang sensitif terhadap penambahan noise ini adalah pada bagian awal.

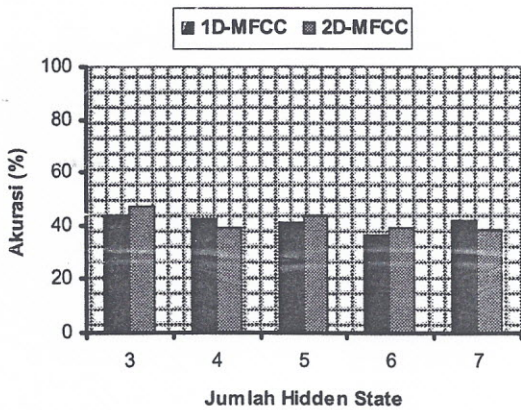
Oleh karena itu dilakukan percobaan kembali dengan tidak memasukkan beberapa komponen ke 1, komponen ke 1 dan ke ke 2, serta komponen ke 1, 2, dan ke 3. Hasil percobaan tersebut seperti diperlihatkan pada Gambar 9.

Terlihat bahwa dengan tidak menyertakan satu koefisien pertama, ternyata teknik 2D-MFCC mampu menaikkan akurasi menjadi 74,8% (naik lebih dari 25%).

Sedangkan untuk 1D-MFCC tetap masih sedikit di atas 50%. Namun demikian, dengan makin meningkatnya level noise, ternyata metode yang diusulkan masih belum memberikan hasil yang memuaskan, seperti diperlihatkan pada Gambar 10.

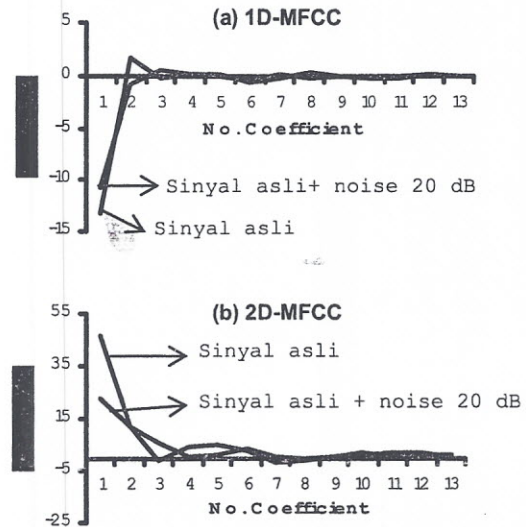


Gambar 6. Perbandingan Akurasi antara 1D-MFCC dengan 2D-MFCC untuk Suara Tanpa Penambahan Noise pada Berbagai Jumlah Hidden State

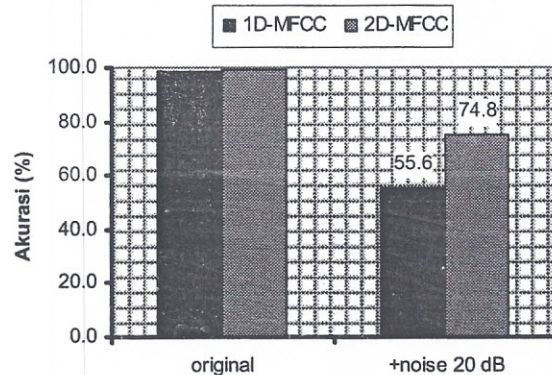


Gambar 7. Perbandingan Akurasi antara 1D-MFCC dengan 2D-MFCC untuk Suara Beroise Noise 20 dB pada Berbagai Jumlah Hidden State

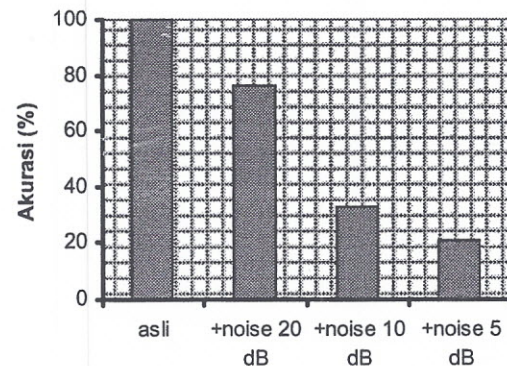
Dibandingkan dengan power spektrum, secara visual, nilai bispektrum sinyal asli dengan sinyal yang sudah ditambah noise masih terlihat mirip, Gambar 11. Hal ini menunjukkan bahwa pemilihan filter yang tepat diharapkan memberikan hasil yang lebih baik.



Gambar 8. Perbandingan Koefisien Vektor Ciri antara Sinyal Asli dengan yang sudah Ditambah Noise 20 dB. 1D-MFCC (a) and 2D-MFCC (b).

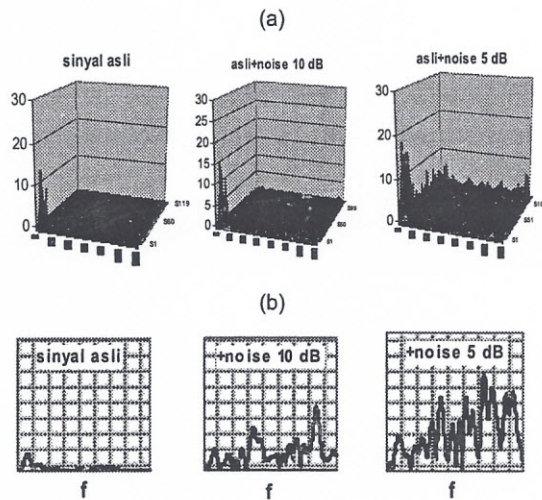


Gambar 9. Perbandingan Akurasi antara 1D-MFCC dengan 2D-MFCC untuk Suara Beroise Noise 20 dB



Gambar 10. Akurasi Sistem Berbasis 2D-MFCC pada Berbagai Noise





Gambar 11. Perbandingan Bispektrum (a) dan Power Spektrum (b) untuk Sinyal Asli, Asli + Noise 10 dB dan Asli + Noise 5 dB

#### 4. Kesimpulan

Dari percobaan yang sudah dilakukan terlihat bahwa perluasan MFCC dari dimensi 1 ke dimensi 2 dapat diterapkan pada sistem identifikasi pembicara. Sistem yang dibangun mampu menghasilkan tingkat pengenalan dengan akurasi sangat baik (> 99%, dan hasil tertinggi mencapai 100%). Untuk data bernoise 20 dB, sistem dengan teknik 2D-MFCC mampu mengenali pembicara dengan akurasi 75%. Sedangkan teknik lama hanya 56%.

#### Ucapan Terima Kasih

Penulis mengucapkan terima kepada Departemen Ilmu Komputer IPB atas ijin penggunaan laboratorium komputer untuk menjalankan percobaan.

#### Daftar Acuan

- [1] T.D. Ganchev, Ph.D. Thesis. Wire Communications Laboratory, Department of Computer and Electrical Engineering, University of Patras, Greece, 2005.
- [2] A. Buono, Sistem Identifikasi Pembicara dengan MFCC sebagai Pengekstraksi Ciri dengan Hidden Markov Model sebagai Classifier, Laporan Teknis Penelitian, Lab. Kecerdasan Komputasional, Fakultas Ilmu Komputer Universitas Indonesia, 2008, tidak dipublikasikan.
- [3] C.L. Nikeas, A.P. Petropulu, Higher Order Spectra Analysis: A Nonlinear Signal Processing Framework, Prentice-Hall, Inc., New Jersey, 1993, p. 14.
- [4] M.I. Fanany, B. Kusumoputro, Thesis Magister, Ilmu Komputer, Fasilkom Universitas Indonesia, Depok, 1998.
- [5] N. Hidayat, B. Kusumoputro, Tesis Magister, Ilmu Komputer, Fasilkom Universitas Indonesia, Depok, 1999.
- [6] A. Triyanto, B. Kusumoputro, Thesis Magister, Ilmu Komputer, Fasilkom Universitas Indonesia, Depok, 2000.
- [7] C. Cornaz, U. Hunkeler, An Automatic Speaker Recognition System, Ecole Polytechnique, Federale De Lausanne, [http://www.ifp.uiuc.edu/~minhdo/teaching/speaker\\_recognition](http://www.ifp.uiuc.edu/~minhdo/teaching/speaker_recognition), 2005.
- [8] L.R. Rabiner, A Tutorial on Hidden Markov Model and Selected Applications in Speech Recognition, Proceeding IEEE 77/2 (1989) 257.
- [9] M. Nilsson, M. Ejnarsson, Master Thesis, Departement of Telecommunications and Signal Processing, Blekinge Institute of Technology, Ronneby, 2002.