Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) DOI: 10.25126/jtiik.202183284
Vol. 8, No. 1, Februari 2021, hlm. 119-126 p-ISSN: 2355-7699
Akreditasi KEMENRISTEKDIKTI, No. 36/E/KPT/2019 e-ISSN: 2528-6579

SISTEM PENGENALAN PEMBICARA DENGAN METODE *WAVELET*-MFCC DAN PENGKLASIFIKASI *HIDDEN MARKOV MODELS* (HMM)

Syahroni Hidayat[*1], Andi Sofyan Anas[2], Siti Agrippina Alodia Yusuf[3], Muhammad Tajuddin[4]

[1,3]Pusat Penelitian dan Pengembangan, Sekawan Institute Nusa Tenggara
[2]Program Studi Rekayasa Perangkat Lunak, Fakultas Teknik dan Desain, Universitas Bumigora
[4]Program Studi Ilmu Komputer, Fakultas Teknik dan Desain, Universitas Bumigora
Email: [1]syahroni.hdyt@gmail.com, [2]andi.sofyan@universitasbumigora.ac.id, [3]alodiaysf@gmail.com, [4]Tajuddin@universitasbumigora.ac.id
[*]Penulis Korespondensi

(Naskah masuk: 19 Februari 2020, diterima untuk diterbitkan: 01 Februari 2021)

Abstrak

Penelitian pengolahan sinyal digital yang berfokus pada pengenalan pembicara telah dimulai sejak beberapa dekade yang lalu, dan telah menghasilkan banyak metode-metode pengenalan pembicara. Di antara algoritma pembentukan koefisien ciri yang telah dikembangkan tersebut, ada dua algoritma yang dapat memberikan akurasi yang tinggi jika diterapkan pada sistem, yaitu *Mel Frequency Cepstral Coefficient* (MFCC) dan *Wavelet*. Penelitian ini bertujuan untuk menguji dan memilih kanal terbaik dari proses *wavelet*-MFCC yang dapat dijadikan sebagai koefisien ciri baru untuk diterapkan pada sistem pengenal pembicara. Koefisien ciri baru tersebut kemudian disebut dengan koefisien ciri *Wavelet*-MFCC. Kofisien ini dibentuk dari merubah kanal hasil dekomposisi *wavelet*, yaitu kanal aproksimasi (cA), kanal detail (cD), dan penggabungannya (cAcD), menjadi koefisien MFCC. Metode dekomposisi *wavelet* yang digunakan adalah metode *dyadic* dengan menerapkan *level* dekomposisi *level* 1 dan *level* 2. Setiap koefisien ciri kemudian menjadi inputan pada sistem pengklasifikasi *Hidden Markov Models* (HMM). Keluaran dari HMM kemudian dihitung akurasinya dan dianalisis. Dari pengujian yang dilakukan, diperoleh bahwa kanal detail (cD) sebagai ciri dapat memberikan akurasi yang sama dengan menggunakan kanal gabungan (cAcD) dan lebih tinggi dari kanal aproksimasi (cA), dengan akurasi sebesar 95%. Hal ini menunjukkan bahwa, kanal detail pada dekomposisi *level* 1 menyimpan ciri suara dari setiap pembicara sehingga sudah cukup untuk dijadikan sebagai koefisien ciri. Maka, penggunaan dekomposisi *level* 1 dan kanal detail cD sebagai ciri *Wavelet-MFCC* pada sistem pengenalan pembicara dapat meringankan dan mempercepat proses komputasi.

Kata kunci: *sistem pengenalan pembicara, Wavelet, MFCC, koefisien Wavelet-MFCC, HMM.*

SPEAKER RECOGNITION SYSTEM USING WAVELET-MFCC METHOD AND HIDDEN MARKOV MODELS (HMM) CLASSIFIER

Abstract

Research in digital signal that focused on speaker recognition has begun since decades ago, and has resulted many speaker recognition methods. there are two algorithms that can provide high accuracy in recognition system, which are Mel Frequency Cepstral Coefficient (MFCC) and Wavelet. the aims of this study is to examine and chose the best channel from wavelet-MFCC process that can be used as new feature coefficient, then called as Wavelet-MFCC features coefficient. The coefficient is built by converting the wavelet decomposition channels, which are approximation (cA), detail (cD), and its combination (cAcD), into the MFCC coefficient. Wavelet dyadic decomposition with level 1 and level 2 of decomposition is applied. Each feature coefficient acts as an input to the HMM classifier. The accuracy of the HMM output is calculated, then analyzed. The obtained results show that the detail chanel (cD) achieve equal accuracy as the combination chanel (cAcD), and higher accuracy compared to aproximation channel (cA), with accuracy 95%. Thus, it can be conclude that the detail channel on level 1 decomposition contains features of each speaker's. Then, cD is enough to be used as a Wavelet-MFCC feature. Thus, its implementation in the SRS can ease and speed up the computing process.

Keywords: *speaker recognition system, Wavelet, MFCC, Wavelet-MFCC coefficient, HMM.*

1. PENDAHULUAN

Sistem pengenalan pembicara merupakan area ilmu pemrosesan sinyal digital yang berhubungan dengan pengenalan pembicara dari suara mereka

(Sujiya dan Chandra, 2017). Penelitian tentang sistem pengenalan pembicara sudah dimulai sejak beberapa dekade yang lalu. Hasilnya ada banyak metode pengenalan pembicara yang telah dikembangkan. Meskipun demikian tujuannya masih tetap sama, yaitu untuk dapat membuat mesin yang dapat digunakan untuk mengekstrak, mengkarakterisasi dan mengidentifikasi pembicara (Todkar et al., 2018). Evaluasi dan pengembangan sistem pengenalan pembicara sampai saat ini tetap dilakukan. NIST mengambil peran tersebut dan yang terbaru evaluasi dan pengembangan sistem pengenalan pembicara tersebut difokuskan pada performa dan penemuan ide baru (Greenberg et al., 2020; Sadjadi et al., 2020). Performa sistem pengenalan pembicara umumnya dipengaruhi, salah satunya, oleh metode ekstraksi ciri. Adapun penemuan ide baru yang digariskan oleh NIST dapat berupa pengembangan metode ekstraksi ciri. Banyak metode ekstraksi ciri yang telah dikembangkan, seperti *MFCC, Wavelet, LPCC*, penggabungan *Wavelet* dan *MFCC,* dan lain-lain (Sharma, Umapathy and Krishnan, 2020).

Dari beberapa metode yang telah dipaparkan, metode yang telah berkembang sangat baik untuk pengenalan pembicara bahkan untuk aplikasi pengenalan suara secara umum adalah algoritma Mel Frequency Cepstral Coefficient (MFCC) (Sharma, Umapathy and Krishnan, 2020; Shirali-Shahreza and Shirali-Shahreza, 2010). Algoritma MFCC dapat membentuk ciri suara yang sangat baik karena meniru pendengaran manusia (Huang, Acero dan Hon, 2001). Algoritma pembentukan ciri lain yang juga terbukti memberikan akurasi sistem pengenalan pembicara setara dengan MFCC adalah *Wavelet*. Bahkan banyak yang mengklaim lebih baik dari MFCC. Hal ini didasarkan pada kemampuan *Wavelet* memetakan sinyal suara ke dalam domain frekuensi-waktu secara bersamaan tanpa terjadinya kehilangan sinyal. Alasan ini menjadi dasar untuk beberapa peneliti menggabungkan MFCC dan *Wavelet* untuk menjadi koefisien ciri (Amelia dan Gunawan, 2019).

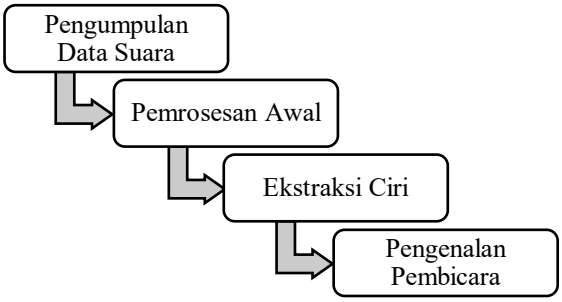
Beberapa peneliti telah melakukan penelitian dengan menggabungkan *Wavelet* dan MFCC. Syahroni, dkk. telah mengidentifikasi bentuk filter MFCC terbaik untuk membentuk koefisien ciri *Wavelet*-MFCC dengan pendekatan algoritma *Mean Best Basis* (MBB). Penerapannya pada sinyal suara vokal dengan dekomposisi *wavelet* paket. Hasilnya membentuk kanal dekomposisi *wavelet* terbaik yang dapat dijadikan sebagai kofisien ciri *wavelet*-MFCC (Hidayat, Abdurahim dan Tajuddin, 2019). Adam dkk telah membangun sistem pengenalan pembicara dengan kata terisolasi dengan pengklasifikasi *Back Propagation Neural Network* (BPNN). Untuk koefisien ciri suara sebagai inputan disebut sebagai *Wavelet Cepstral Coefficient* (WCC). WCC dibentuk dari seluruh kanal hasil dekomposisi *wavelet dyadic*(koefisien aproksimasi, cA, dan koefisien detail, cD) yang diaplikasikan algoritma *log power spectrum* dan DCT terhadapnya (Adam, Salam dan Gunawan, 2013).

Serupa dengan apa yang dilakukan Adam, dkk., Amelia dan Gunawan membentuk koefisien ciri gabungan *wavelet*-MFCC yang disebut sebagai DWT-MFCC. Perbedaannya terletak pada kanal yang dijadikan sebagai koefisien ciri, yaitu pada penelitian ini hanya mengubah kanal koefisien aproksimasi cA saja (Amelia dan Gunawan, 2019).

Dari dua penelitian terakhir yang telah disebutkan, Adam dkk. telah menggunakan gabungan koefisien aproksimasi cA dan koefisien detail cD sebagai ciri utama, sedangkan Amelia dkk. hanya menggunakan cA sebagai ciri utama. Penelitian ini berfokus pada pengujian seluruh kanal hasil dekomposisi *wavelet* agar dapat dijadikan sebagai koefisien ciri *wavelet*-MFCC baru. Algoritma pengklasifikasi yang digunakan adalah *Hidden Markov Models* (HMM). Kemudian setiap koefisien ciri akan dianalisis akurasinya. Selain itu, akan dianalisis juga pengaruh jenis kelamin pembicara terhadap penggunaan setiap koefisien ciri *wavelet*-MFCC.

2. METODE PENELITIAN

Secara umum tahapan yang dilalui dalam penelitian ini ditunjukkan pada Gambar 1. Tahapan-tahapannya adalah pengumpulan data, pemrosesan awal, proses ekstraksi ciri, dan pengenalan pembicara.



Gambar 1. Tahapan penelitian

2.1 Pengumpulan Data Suara

Penelitian diawali dengan proses perekaman data suara. Data suara yang direkam adalah ucapan kata 'HADIR'. Sampel suara diperoleh dari 30 orang pembicara dewasa, masing-masing 20 orang pembicara laki-laki dan 10 orang pembicara perempuan. Para pembicara mengucapkan kata hadir dengan 10 kali perulangan. Sehingga total *dataset* suara yang diperoleh sebanyak 300 data suara. Dari seluruh data tersebut sebanyak 80% dari suara rekaman setiap pembicara dijadikan sebagai data pelatihan dan sisanya sebesar 20% sebagai data pengujian. Jumlah sampel pembicara laki-laki dan perempuan yang tidak seimbang ini didasarkan pada performa suara perempuan yang lebih baik dalam

sistem pengenalan suara secara umum (Sawalha and Abushariah, 2013). Hal ini didasarkan pada karakteristik dasar suara perempuan yang memiliki *'loudness'* yang lebih tinggi dari suara laki-laki apalagi jika diterapkan ekstraksi ciri *wavelet*-MFCC terhadapnya (Mason and Thompson, 1993), sehingga diasumsikan jumlah pembicara perempuan dapat mempengaruhi performa sistem secara keseluruhan.

Alat yang digunakan untuk merekam suara adalah sebuah *headphone* yang terintegrasi dengan mikrofon. Sementara aplikasi perekaman yang digunakan adalah *Audacity*. Ada dua jenis pengaturan yang dilakukan dalam proses perekaman, pertama adalah pengaturan kondisi lingkungan perekaman dan kedua pengaturan properti perekaman dalam aplikasi *Audacity*. Pada pengaturan pertama, ruangan yang digunakan adalah ruangan tertutup. Hal ini ditujukan untuk mengurangi adanya gangguan (*noise*) yang ikut terekam. Adapun alat perekaman diletakkan dengan jarak 0.5 meter dari pembicara. Untuk pengaturan kedua, properti perekaman, mengikuti Tabel 1 berikut:

Tabel 1. Properti Perekaman

No	Variabel	Nilai
1	Channel Perekaman	Mono
2	Frekuensi Sampling	8096 Hz
3	BitsPerSample	16-bit PCM
4	Ekstensi penyimpanan	*.wav

2.2 Pemrosesan Awal

Pemrosesan awal ini bertujuan untuk meningkatkan kualitas suara hasil perekaman. Ada tiga tahapan yang dilakukan yaitu, mengurangi gangguan (*noise*), deteksi aktivitas suara, dan normalisasi sinyal suara. Pada tahap pertama diterapkan *filter pre emphasis*. *Filter* ini digunakan untuk mengurangi gangguan (*noise*) sekaligus meningkatkan sinyal berfrekuensi tinggi dalam rekaman suara tersebut (Huang et al., 2001). Pada prinsipnya nilai koefisien *filter pre emphasis* berikisar antara 0.9 – 0.97. Pada penelitian ini digunakan nilai koefisien filter α = 0.97. Formula untuk menghitung *filter pre emphasis* ditunjukkan pada persamaan (1). y[n] adalah hasil filter, x[n] adalah sinyal ke-n, dan x[n-1] adalah sinyal sebelumnya.

$$y[n] = x[n] - \alpha * x[n-1] \quad (1)$$

Setelah dilakukan penghilangan *noise*, diterapkan algoritma deteksi aktifitas suara (*Voice Activity Detection*). Algoritma ini ditujukan untuk menghilangkan sinyal *silence* sehingga hanya sinyal suara saja yang akan diolah pada proses berikutnya. Umumnya algoritma ini menggunakan nilai energi, *E*, sebagai nilai ambang batas untuk membedakan suara dengan *silence* (Asliyan, 2011). Rumus energi dalam sinyal suara ditunjukkan pada persamaan (2).

$$E = \frac{1}{N}\sum_{n=1}^{N}|x(n)|^2 \quad (2)$$

Terakhir adalah normalisasi, S_{norm}. Normalisasi dilakukan untuk mendapatkan nilai *magnitude* yang seragam pada seluruh *dataset* sinyal suara. Proses ini menghasilkan sinyal suara dengan nilai *magnitude* maksimal ± 1 (Hidayat, Hidayat and Adji, 2015). Persamaan (3) digunakan untuk memperoleh sinyal suara ternormalisasi. S_{norm} adalah sinyal ternormalisasi dan $max|S|$ adalah nilai maksimum sinyal tersebut.

$$S_{Norm} = \frac{S}{max|S|} \quad (3)$$

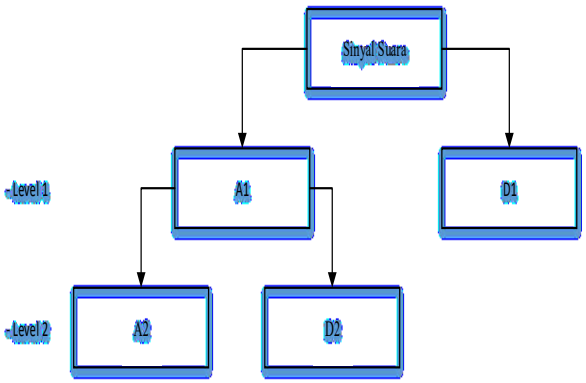
2.3 Ekstraksi Ciri

Proses ekstraksi ciri merupakan proses untuk mengambil fitur ciri dari sinyal suara rekaman. Ciri yang diambil harus unik dan memiliki variabilitas yang tinggi dengan ciri yang lainnya. Pada tahap ini diterapkan dua metode yang digabungkan menjadi satu yaitu *wavelet* dan MFCC. Penggabungan ini dilakukan untuk memanfaatkan kelebihan-kelebihan yang dimiliki oleh kedua metode tersebut (Hidayat, Priyatmadi dan Ikawijaya, 2015; Hidayat, Hidayat dan Adji, 2015).

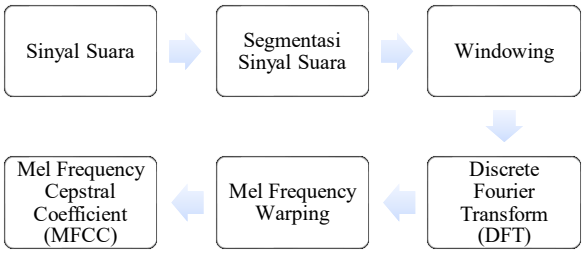
Pada metode ekstraksi fitur *wavelet* digunakan metode dekomposisi *dyadic*. Proses dekomposisi *wavelet dyadic* ditunjukkan pada Gambar 2. Jenis *family wavelet* yang digunakan adalah Haar *wavelet*. *Wavelet* ini adalah *wavelet* paling sederhana dan merupakan induk untuk pengembangan jenis *family wavelet* lainnya. *Level* dekomposisi yang diterapkan adalah *level* 1 dan *level* 2. Pada setiap *level* dekomposisi akan membagi sinyal menjadi kanal Aproksimasi (A) dan Detail (D). Kanal A merupakan hasil dekomposisi setelah diterapkan *filter* lolos rendah g[n] terhadapnya, sedangkan kanal D merupakan hasil dekomposisi setelah diterapkannya *filter* lolos tinggi h[n]. Sehingga pada kanal A berisi sinyal suara berfrekuensi rendah sedangkan pada kanal D berisi sinyal suara berfrekuensi tinggi. *Filter* lolos rendah dan *filter* lolos tinggi yang diimplementasikan pada proses dekomposisi *wavelet* dinyatakan oleh persamaan (4) dan (5). Setiap koefisien yang terdapat pada kanal hasil dekomposisi *wavelet* akan diubah menjadi koefisien MFCC (Hidayat, Abdurahim and Tajuddin, 2019).

$$A[n] = \sum_{k=-\infty}^{\infty} s[k]g[2n-k] \quad (4)$$

$$D[n] = \sum_{k=-\infty}^{\infty} s[k]h[2n-k] \quad (5)$$



Gambar 2. Dekomposisi *wavelet level* 2 metode *dyadic*



Gambar 3. Tahapan proses MFCC

MFCC adalah koefisien ciri yang diperoleh dengan memanfaatkan transformasi *fourier* dalam prosesnya (Huang, Acero dan Hon, 2001). Tahapan untuk memperoleh ciri MFCC ditunjukkan pada Gambar 3.

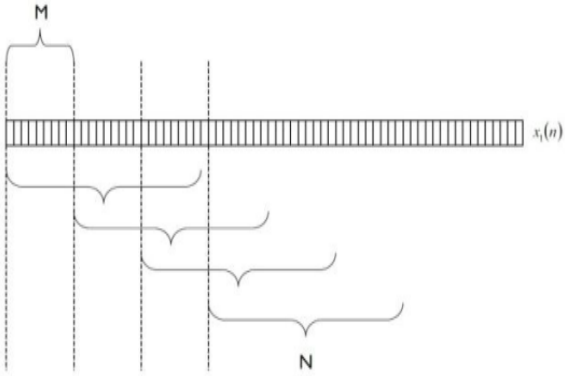
Pertama-tama sinyal suara disegmentasi dengan ukuran N = 0.025 detik. Pergeseran segmentasinya sebesar M = 0.01 detik. Gambar 4 menunjukkan cara kerja proses segmentasi. Setiap segmen kemudian diterapkan kepadanya proses *windowing*. Tipe *window* $w(n)$ yang digunakan adalah *window* Hamming. *Window* Hamming diperoleh dengan mengimplementasikan persamaan (6).

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right) \quad (6)$$

dengan $0 \leq n \leq N$. N adalah panjang segmen (Jurafsky and Martin, 2008).

Selanjutnya adalah proses transformasi *fourier*. Dalam hal ini diterapkan algoritma transformasi *fourier* diskrit. Jumlah titik yang digunakan pada algoritma DFT adalah sebanyak 512 titik. Algoritma ini dinyatakan pada persamaan (7) berikut. $X[k]$ adalah hasil DFT dan $x[n]$ adalah sinyal diskrit ke-n.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N} \quad (7)$$



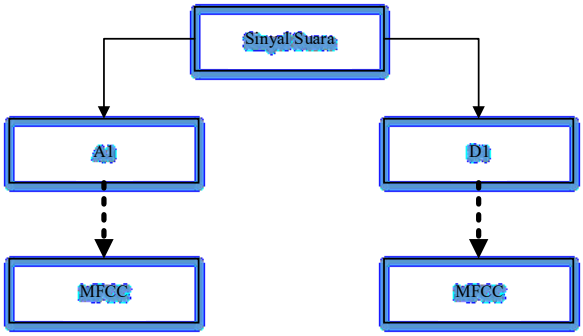
Gambar 4. Tahapan proses segmentasi sinyal

Setiap segmen yang telah ditransformasi fourier kemudian dibungkus nilai spektrumnya dengan menggunakan *filter* segitiga. Proses ini disebut sebagai proses *Mel Frequency Warping*. Hal ini diaplikasikan karena isyarat suara berbeda dengan persepsi pendengaran manusia, dimana isyarat suara tidaklah memiliki frekuensi dengan skala yang linier. Untuk memperoleh nilai frekuensi mel, *mel*(f) dan filter segitiga $H_m[k]$ digunakan persamaan (8) dan (9). Dalam penelitian ini jumlah filter segitiga yang dibentuk adalah 26. Artinya setelah proses ini akan terbentuk nilai koefisien mel sebanyak 26. Kemudian dari 26 nilai koefisien ini hanya 13 nilai koefisien pertama yang digunakan sebagai ciri MFCC (Hossan, Memon and Gregory, 2010).

$$mel(f) = 1125\ln(1 + \frac{f_{hz}}{700}) \quad (8)$$

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \dfrac{(k - f[m-1])}{(f[m] - f[m-1])} & f[m-1] \leq k \leq f[m] \\ \dfrac{(f[m+1] - k)}{(f\{m+1] - f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1 \end{cases} \quad (9)$$

Lebih ringkas, gambaran proses pembentukan koefisien ciri *Wavelet*-MFCC ditunjukkan pada Gambar 5.

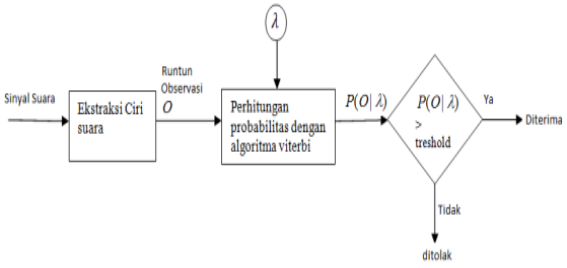


Gambar 5. Proses ekstraksi ciri *wavelet*-MFCC *level* 1

Dari 240 data *training* yang diolah kemudian dibuat 30 vektor referensi yang mewakili setiap pembicara. Setiap vektor berisi nilai koefisien mel dengan panjang 26 data, 13 data pertama berisi koefisien mel yang diperoleh dari hasil dekomposisi *level* 1 kanal aproksimasi (A1), sedangkan 13 data selanjutnya diperoleh dari kanal detail (D1).

2.4 Pengenalan Pembicara

Terdapat beberapa algoritma yang dapat diimplementasikan untuk melakukan pengenalan/verifikasi pembicara. Salah satunya yang paling umum dan terbukti memiliki performansi yang sangat baik adalah metode *Hidden Markov Model* (HMM). Secara umum, gambaran sistem pengenalan pembicara menggunakan HMM ditunjukkan pada Gambar 6 (Darmawan dan Ariessaputra, 2018). Hasil ekstraksi ciri suara beperan sebagai inputan pada sistem HMM, disebut sebagai runtun observasi (O). Setiap runtun observasi ini kemudian ditentukan nilai probabilitas observasinya $b_j(O_t)$ dengan persamaan (10). Hasilnya berupa model HMM untuk setiap pembicara (λ). Stelah model HMM diperoleh dihitung probabilitas model $P(O|\lambda)$ dengan algoritma Viterbi. Hasil dari perhitungan probabilitas model ini dijadikan sebagai nilai similaritas.



Gambar 6. Sistem pengenalan pembicara HMM

$$b_j(O_t) = \frac{1}{1+d(O_t,\mu_j)} \quad (10)$$

Dengan $d(O_t, \mu_j)$ adalah jarak Euclidean yang diukur dengan persamaan (11).

$$d(O_t, \mu_j) = \sqrt{\sum_{k=1}^{M}(O_{tk} - \mu_{jk})^2} \quad (11)$$

Selanjutnya untuk mengukur akurasi sistem digunakan persamaan (12) berikut ini:

$$Akurasi = \frac{Jumlah\ pembicara\ dikenali\ benar}{Jumlah\ total\ data\ yang\ diuji} x100\% \quad (12)$$

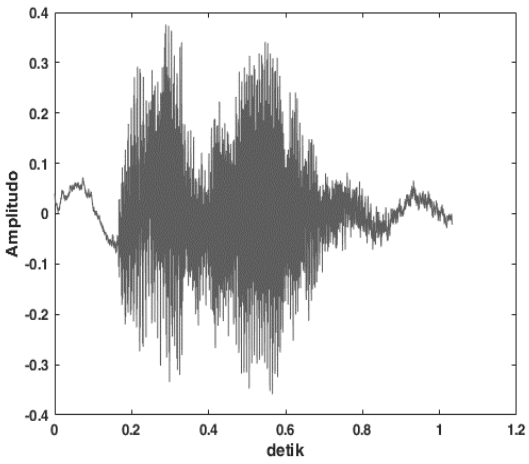
Penelitian ini difokuskan pada tiga kondisi berikut, yaitu (1) penerapan koefisien aproksimasi cA sebagai koefisien ciri *Wavelet*-MFCC, (2) penerapan koefisien detail cD sebagai koefisien ciri *Wavelet*-MFCC, dan (3) penerapan dari penggabungan kedua koefisien cA+cD sebagai koefisien ciri *Wavelet*-MFCC. Penggabungan yang dilakukan berupa konkatenasi antara kedua ciri cA dan cD.

3. HASIL DAN PEMBAHASAN

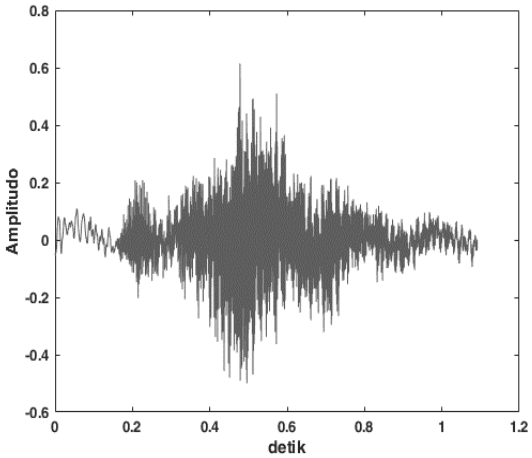
Pada bagian ini dibahas tentang hasil perekaman, hasil peningkatan kualitas suara, hasil ekstraksi ciri, dan hasil pengujian sistem berikut dengan pembahasannya.

3.1. Hasil Perekaman

Hasil perekaman ditunjukkan pada gambar 7 dan gambar 8. Masing-masing secara berurutan mewakili sampel suara pembicara pria dan wanita. Dari hasil perekaman terlihat durasi perekaman kata hadir rata-rata lebih dari 1 detik. Pada awal perekaman terlihat bahwa bentuk sinyal suara yang dihasilkan tidak tepat di angka 0. Ini menunjukkan bahwa perlu diimplementasikan *filter pre-emphasis* pada data suara tersebut.



Gambar 7. hasil perekaman suara pria



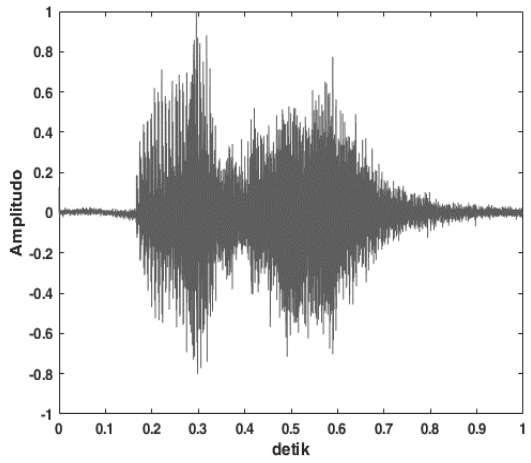
Gambar 8. hasil perekaman suara wanita

Apabila dilihat dari nilai maksimum amplitudonya, terdapat perbedaan yang sangat jelas. Nilai amplitudo untuk sampel suara pria bernilai < 0.4

sementara nilai amplitudo suara wanita mencapai > 0.6. Perbedaan nilai amplitudo ini melatarbelakangi pentingnya dilakukan normalisasi sinyal suara.

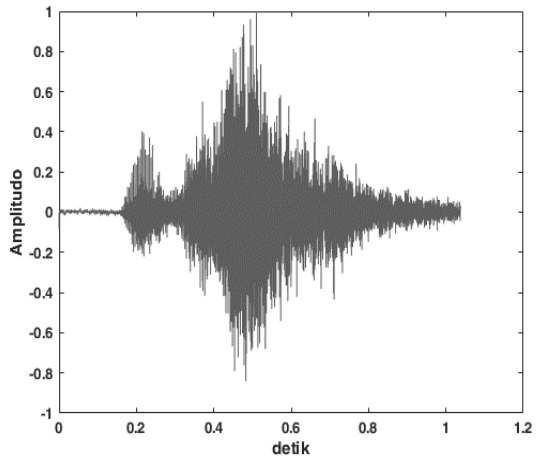
3.2. Hasil Peningkatan Kualitas Suara

Hasil peningkatan kualitas yang meliputi beberapa langkah seperti yang telah dipaparkan pada bagian metode penelitian ditunjukkan pada Gambar 9 dan Gambar 10. Penerapan *filter pre emphasis* dapat dilihat dari bentuk sinyal suara pada bagian awalnya yang sudah tepat berada pada angka 0. Sementara untuk hasil proses normalisasi ditunjukkan oleh besar nilai amplitudo maksimal untuk kedua sinyal yang saat ini telah bernilai 1. Normalisasi ini hanya menyamaratakan nilai puncak dan tidak mempengaruhi informasi yang dikandung oleh suara rekaman asli. Dengan demikian, seluruh sinyal suara yang akan diolah selanjutnya masih asli dan sudah pasti dalam bentuk normal.



Gambar 9. Hasil peningkatan kualitas suara pria

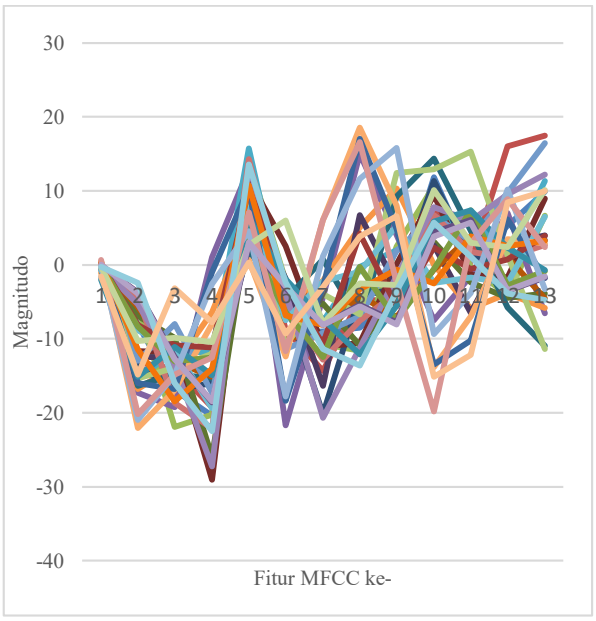
Berikutnya hasil dari proses deteksi aktifitas suara hanya dapat dilihat jelas jika membandingkan Gambar 7 dengan Gambar 9. Dapat dilihat dengan jelas durasi suara pada Gambar 9 sama dengan 1 detik setelah dihilangkannya sinyal yang berkategori sebagai sinyal *silence*.



Gambar 10. Hasil peningkatan kualitas suara wanita

3.3 Hasil Ekstraksi Ciri

Hasil ekstraksi ciri *Wavelet*-MFCC ditunjukkan pada Gambar 11. Sumbu horizontal menyatakan jumlah fitur MFCC sebanyak 13 buah, sedangkan sumbu vertikal menyatakan nilai magnitudo untuk setiap fitur. Pada ciri cA atau cD hanya terdapat 13 fitur untuk setiap pembicara, sedangkan pada ciri gabungannya terdapat 26 fitur ciri untuk setiap pembicara. Total terdapat 30 grafik yang masing-masing mewakili ciri setiap pembicara. Ciri ini merupakan ciri masukan (runtun observasi, O) yang akan diolah oleh algoritma HMM untuk membentuk model HMM, λ, bagi setiap pembicara. Dapat diperhatikan bahwa setiap ciri yang telah diekstrak memiliki tingkat variabilitas yang masih rendah terhadap ciri lainnya. Nilai ini akan ditingkatkan dengan menghitung nilai probabilitas modelnya, $P(O|\lambda)$, dengan algoritma Viterbi. Setelah itu, model yang diperoleh dari penerapan algoritma Viterbi disimpan sebagai ciri referensi.



Gambar 11. Hasil ekstraksi ciri cA1

3.4 Hasil Pengujian Sistem

Hasil pengujian sistem untuk tiga kondisi yang telah disebutkan pada bagian metode penelitian ditunjukkan pada Tabel 2. Hasil pengujian dengan menggunakan kanal cA sebagai ciri *Wavelet*-MFCC menunjukkan bahwa pada dekomposisi *level*-1 akurasi pengenalan pembicaranya sebesar 91.7 %. Sedangkan pada dekomposisi *level*-2 akurasinya sebesar 90%. Sedangkan untuk percobaan menggunakan kanal cD sebagai ciri *wavelet*-MFCC diperoleh akurasi pada dekomposisi *level* 1 dan *level* 2 secara berturut-turut adalah 95% dan 78.3%. Adapun dengan menggunakan ciri gabungan

diperoleh akurasi untuk masing-masing *level* dekomposisi sebesar 95% dan 85%.

Tabel 2. Akurasi sistem pengenalan pembicara

Koefisien	Dekomposisi	
	level 1	*level* 2
cA	91.7%	90.0%
cD	95.0%	78.3%
cA + cD	95.0%	85.0%

Secara global dapat dilihat bahwa penambahan *level* dekomposisi *wavelet* dapat memberikan penurunan akurasi. Penurunan tingkat akurasi yang paling signifikan ditunjukkan jika menggunakan kanal detail cD dan kanal gabungan sebagai koefisien ciri *wavelet*-MFCC. Sedangkan jika menggunakan kanal aproksimasi cA penurunan akurasi tidak signifikan. Hal ini dapat disebabkan karena pada kanal aproksimasi tersimpan nilai koefisien sinyal yang mendekati nilai asli dari sinyal tersebut. Sedangkan koefisien cD menyimpan koefisien detail dari sinyal suara.

Meskipun demikian, perhatian yang paling besar pada hasil penelitian ini dapat dilimpahkan pada hasil penggunaan kanal cD sebagai ciri pada dekomposisi *level* 1. Tingkat akurasi yang lebih tinggi dari penerapan kanal cA dan tingkat akurasi yang sama dengan jika menggunakan ciri gabungan menunjukkan bahwa penggunaan hanya koefisien detail cD dapat menjadi alternatif baru dalam pembentukan koefisien ciri, khususnya pada sistem pengenalan pembicara.

Kondisi ini merupakan kondisi yang langka, karena umumnya untuk penelitian-penelitian tentang pengenalan atau verifikasi pembicara yang menggunakan metode *wavelet* seringkali menggunakan seluruh hasil dekomposisi sebagai ciri. Namun ternyata, penggunaan koefisien detail saja yang dikombinasikan dengan perhitungan HMM sudah dapat memberikan akurasi pengenalan sebesar 95%. Sehingga, dalam kasus pengenalan pembicara dapat dikatakan bahwa koefisien detail *wavelet* mengandung benyak informasi yang mencirikan pembicara dibandingkan dengan koefisien aproksimasinya. Selain itu, tentunya penggunaan hanya koefisein detail cD pada *level* dekomposisi ke-1 sebagai koefisien ciri *Wavelet*-MFCC tentunya dapat menghemat proses komputasi.

Performa sistem pengenalan pembicara penelitian ini jika dilihat berdasarkan pembicara ditunjukkan pada Tabel 3. Hanya penggunaan ciri *wavelet*-MFCC dengan dekomposisi *level* 1 saja yang ditampilkan. Ini didasarkan pada performa terbaik yang telah ditampilkan pada Tabel 2. Dapat dilihat bahwa performa sistem untuk pembicara laki-laki lebih tinggi dari pembicara perempuan. Meskipun demikian, perbedaannya hanya sebesar 2.5 %. Jika dilihat dari kanal yang digunakan sebagai ciri, tidak terjadi perubahan signifikan. Pada ciri suara laki-laki, kanal cD masih dapat memberikan performa yang sama dengan kanal gabungan. Sedangkan pada kanal cD untuk ciri suara perempuan, memberikan performa yang lebih tinggi dari kanal cA maupun kanal gabungan.

Tabel 3. Akurasi sistem pengenalan pembicara berdasarkan jenis kelamin pada dekomposisi *level* 1 *wavelet* Haar

Jenis Kelamin	cA1	cD1	cA1+cD1
Laki-laki	92.5 %	97.5 %	97.5 %
Perempuan	90.0 %	95.0 %	90.0 %

Hasil ini menunjukkan dan menguatkan bahwa penggunaan kanal cD saja sebagai *wavelet*-MFCC untuk sistem pengenal pembicara sudah sangat tepat. Akhirnya dapat disimpulkan juga bahwa kanal detail cD dapat mengekstrak dengan baik ciri khas suara dari masing-masing pembicara.

4. KESIMPULAN

Dari hasil penelitian yang telah dipaparkan di atas dapat disimpulkan bahwa penggunaan kanal koefisien detail, cD, sebagai basis pembentukan ciri MFCC memberikan hasil yang paling baik. Bahkan dilihat dari pengaruh jenis kelamin pembicara, kanal detail dapat mengekstraksi ciri khas suara dari setiap pembicara. Hasil ini baru terbukti untuk aplikasi pengenalan atau verifikasi pembicara. Sehingga pada penelitian selanjutnya dapat difokuskan untuk menerapkan metode ini pada aplikasi pengenalan suara. Selain itu penelitian juga dapat dilanjutkan untuk meningkatkan akurasi sistem dengan memodifikasi bagian pengenalan suara menggunakan algoritma yang lebih kompleks dan terbaru seperti *Deep Learning*.

DAFTAR PUSTAKA

ADAM, T.B., SALAM, M.S. & GUNAWAN, T.S., 2013. *Wavelet* Cesptral Coefficients For Isolated Speech Recognition. *Telkomnika*, 11(5), Pp.2731–2738.

AMELIA, F. & GUNAWAN, D., 2019. Dwt-Mfcc Method For Speaker Recognition System With Noise. *2019 7th International Conference On Smart Computing And Communications, Icscc 2019*, Pp.1–5.

ASLIYAN, R., 2011. Syllable Based Speech Recognition. In: I. Ipsic, Ed. *Speech Technologies*. [Online] Intech.Pp.263–284. Available At: <Http://Www.Intechopen.Com/Books/Speech-Technologies/Syllable-Based-Speech-Recognition>.

DARMAWAN, B. & ARIESSAPUTRA, S., 2018. Sistem Pengenalan Dan Verifikasi Pembicara Hmm. In: *Citee*. Pp.68–73.

GREENBERG, C.S., MASON, L.P., SADJADI, S.O.

& REYNOLDS, D.A., 2020. Two Decades Of Speaker Recognition Evaluation At The National Institute Of Standards And Technology. *Computer Speech And Language*, [Online] 60, P.101032. Available At: <Https://Doi.Org/10.1016/J.Csl.2019.101032>.

HIDAYAT, R., PRIYATMADI & IKAWIJAYA, W., 2015. *Wavelet* Based Feature Extraction For The Vowel Sound. In: *2015 International Conference On Information Technology Systems And Innovation, Icitsi 2015 - Proceedings*. Pp.1–4.

HIDAYAT, S., ABDURAHIM & TAJUDDIN, M., 2019. Evaluation And Design Of *Wavelet* Packet Cepstral Coefficient (Wpcc) For A Noisy Indonesian Vowels Signal. *Journal Of Physics: Conference Series Paper*, 1211(012023).

HIDAYAT, S., HIDAYAT, R. & ADJI, T.B., 2015. Speech Recognition Of Kv-Patterned Indonesian Syllable Using Mfcc, *Wavelet* And Hmm. *Jurnal Ilmiah Kursor*, 8(2), Pp.67–78.

HOSSAN, M.A., MEMON, S. & GREGORY, M.A., 2010. A Novel Approach For Mfcc Feature Extraction. *4th International Conference On Signal Processing And Communication Systems, Icspcs'2010 - Proceedings*.

HUANG, X., ACERO, A., HON, H.-W. & REDDY, R., 2001. *Spoken Language Processing: A Guide To Theory, Algorithm And System Development*. United States: Prentice Hall Ptr.

JURAFSKY, D. & MARTIN, J.H., 2008. *Speech And Language Processing: An Introduction To Natural Language Processing, Computational Linguistics, And Speech Recognition*. 1st Ed. Prentice Hall.

MASON, J.S. & THOMPSON, J., 1993. Gender Effects In Speaker Recognition. In: *Proc. Icsp-93*. Pp.733–736.

SADJADI, S.O., GREENBERG, C., SINGER, E., REYNOLDS, D., MASON, L. & HERNANDEZ-CORDERO, J., 2020. The 2019 Nist Speaker Recognition Evaluation Cts Challenge. Pp.266–272.

SAWALHA, M. & ABUSHARIAH, M.A.M., 2013. The Effects Of Speakers ' Gender , Age , And Region On Overall Performance Of Arabic Automatic Speech Recognition Systems Using The Phonetically Rich And Balanced Modern Standard Arabic Speech Corpus. In: *Proceedings Of The 2nd Workshop Of Arabic Corpus Linguistics Wacl-2*.

SHARMA, G., UMAPATHY, K. & KRISHNAN, S., 2020. Trends In Audio Signal Feature Extraction Methods. *Applied Acoustics*, [Online] 158, P.107020. Available At: <Https://Doi.Org/10.1016/J.Apacoust.2019.107020>.

SHIRALI-SHAHREZA, M.H. & SHIRALI-SHAHREZA, S., 2010. Effect Of Mfcc Normalization On Vector Quantization Based Speaker Identification. In: *2010 Ieee International Symposium On Signal Processing And Information Technology, Isspit 2010*. Pp.250–253.

SUJIYA, S. & CHANDRA, E., 2017. A Review On Speaker Recognition. *International Journal Of Engineering And Technology (Ijet)*, 9(3), Pp.1592–1598.

TODKAR, S.P., BABAR, S.S., AMBIKE, R.U., SURYAKAR, P.B. & PRASAD, J.R., 2018. Speaker Recognition Techniques: A Review. In: *2018 3rd International Conference For Convergence In Technology, I2ct 2018*. Ieee.Pp.1–5.