Modifikasi Metode MFCC untuk Identifikasi Pembicara di Lingkungan Ber-*Noise*

Yanuar Risah Prayogi
yanuar.tif@unusida.ac.id

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Nahdlatul Ulama Sidoarjo

Abstract*—Some feature extraction methods for speaker identification systems have a disadvantage that their accuracy decreases when the environment is noisy.The Mel Frequency Cepstral Coefficient (MFCC) method is a noise-sensitive voice extraction method.The MFCC method produces high accuracy when in a clean environment. Conversely, when in a noise environment, accuracy drops dramatically.This research proposes feature extraction methods using MFCC combined with endpoint detection algorithm.The endpoint detection algorithm separates speech and nonspeech regions.Nonspeech area usually contains more noise so it can be used as noise information.The noise information is extracted and generates a noise frequency magnitude.Test of the proposed method yields higher accuracy values for all noise types and SNR levels.The accuracy produced by the proposed method is 14.69% higher than the MFCC method, 6.4% compared to the MFCC + wiener method, and 2.74% compared to the MFCC + Spectral Subtraction (SS) method.

Intisari— Beberapa metode ekstraksi fitur untuk sistem identifikasi pembicara memiliki kelemahan yaitu ketika dilingkungan berderau hasil akurasinya menurun. Metode ekstraksi fitur*Mel-Frequency Cepstral Coefficient* (MFCC) merupakan metode ekstraksi sinyal suara yang peka terhadap derau. Metode MFCC menghasilkan akurasi yang tinggi ketika dilingkungan yang bersih. Sebaliknya ketika dilingkungan yang berderau akurasi yang dihasilkan turun drastis. Penelitian ini mengusulkan metode ekstraksi fitur menggunakan MFCC digabung dengan algoritma deteksi endpoint. Algoritma deteksi endpoint memisahkan daerah *speech* dan *nonspeech*. Daerah *nonspeech* biasanya lebih banyak berisi derau sehingga bisa dijadikan informasi derau. Informasi derau diekstrak dan menghasilkan magnitude frekuensi derau. Uji coba metode yang diusulkan menghasilkan nilai akurasi yang lebih tinggi pada semua tipe derau dan tingkat SNR. Akurasi yang dihasilkan oleh metode yang diusulkan lebih tinggi 14.69% dibanding metode MFCC, 6.4% dibanding metode MFCC+wiener, dan 2.74% dibanding metode MFCC+Spectral Subtraction (SS).

Kata Kunci—identifikasi pembicara, Mel Frequency Cepstral Coefficient (MFCC) termodifikasi.

I. PENDAHULUAN

Identifikasi pembicara adalah proses mencocokkan identitas pembicara menggunakan suara pengguna[6]. Pada sistem identifikasi pembicara, database sudah berisi daftar identitas pembicara. Pembicara mengklaim sebuah identitas menggunakan suaranya. Sedangkan sistem mencocokkan suara pembicara dengan database yang ada pada sistem. Ada baberapa kelemahan pada sistem identifikasi pembicara salah satunya adalah tidak bisa akurat ketika di lingkungan yang berderau [13]. Nilai akurasi dari sistem identifikasi pembicara mengalami penurunan secara drastis ketika di lingkungan sebenarnya. Penyebabnya adalah lingkungan ketika pelatihan dan pengenalan berbeda jauh. Derau yang ada di lingkungan sebenarnya atau ketika pengujian ternyata lebih banyak sehingga akurasi dari sistem identifikasi pembicara rendah.

Metode ekstraksi sinyal suara yang umum,banyakdigunakan, dan terkenal adalah MFCC[3]. Cepstral coefficient yang dihasilkan peka terhadap derau [13]. Metode MFCC menghasilkan akurasi yang sangat tinggi ketika di tempat yang tidak berderau. Sedangkan di lingkungan yang berderau, akurasi metode MFCC turun drastis. Sehingga diperlukan modifikasi metode MFCC agar bisa menghasilkan akurasi yang tinggi baik di kondisi tanpa derau maupun dengan derau.Dari paparan penelitian sebelumnya, metode MFCC masih perlu ditingkatkan kinerjanya untuk lingkungan bersih maupun berderau. MFCC perlu dimodifikasi untuk meningkatkan akurasi baik di lingkungan tingkat SNR tinggi maupun rendah. Modifikasi harus mampu mengurangi bahkan menghilangkan derau pada tingkat derau yang tinggi maupun rendah.

Pada penelitian ini, penulis mengusulkan untuk memodifikasi metode MFCC dengan menambahkan proses ekstrasi informasi dari sinyal hasil algoritma deteksi *endpoint*. Algoritma tersebut mengolah sinyal suara sehingga area *nonspeech* dan *speech*terpisah [11]. *Nonspeech* tidak digunakan untuk proses selanjutnya karena tidak ada informasi yang bisa diekstrak sehingga bisa disebut sebagai residu. Sedangkan pada sinyal berderau, daerah *nonspeech* diisi oleh informasi derau sehingga bisa dimanfaatkan dan diekstrak untuk mendapatkan infomasi derau [9].Daerah *nonspeech* atau sisa suara diekstrak dan diubah ke domain frekuensi. Hasil dari ekstraksi *nonspeech* adalah magnitude frekuensi derau.

Informasi derau tadi digunakan untuk menghilangkan derau pada daerah *speech*. Paresh dan Nikita melakukan proses pengurangan terjadi di domain frekuensi[1].

II. METODE PENELITIAN

Sistem yang akan dibangun adalah identifikasi pembicara menggunakan modifikasi metode MFCC. *Metode Mel-Frequency Cepstral Coefficient* (MFCC) mengalami modifikasi pada bagian setelah FFT dan sebelum *Mel-frequency Filtering*. Pada bagian ini terdapat penambahan satu blok proses yang digunakan untuk mengurangi sinyal utama dengan sinyal informasi sinyal derau [8]. Sedangkan pada bagian awal terdapat proses deteksi endpoint yang digunakan untuk memisahkan sinyal utama dan residu sinyal suara. Gambar proses dari metode MFCC yang termodifikasi ditunjukkan pada Gambar 1.

Penelitian sebelumnya yang dilakukan oleh Lim & Oppenheim yaitu Spectral Subtraction (SS) [7]dan penelitian Paresh dan Nikita [1]menjadi dasar dari modifikasi metode MFCC. Metode *Spectral Subtraction* mengekstrak *frame* awal dari sinyal suara untuk mengurangi bahkan menghilangkan derau. Sedangkan Paresh dan Nikita menghilangkan frekuensi derau menggunakan *wiener filter*.

Pada Gambar 1, blok proses yang berwarna abu-abu adalah blok proses yang digunakan untuk mengekstrak informasi sinyal derau. Metode *Spectral Subtraction* menggunakan frame sebelumnya sebagai informasi. Metode yang diusulkan menggunakan *frame-frame* sisa deteksi *endpoint* untuk diekstrak [4]. Sinyal utama dan sinyal sisa hasil deteksi *endpoint* sama-sama diekstrak sehingga dihasilkan magnitude frekuensi. Kemudian magnitude sinyal utama dikurangi magnitude sinyal sisa [5].

A. *Praproses (Deteksi Endpoint)*

Pada praproses terdapat deteksi *endpoint* yang digunakan untuk memisahkan antara sinyal utama dengan sinyal derau. Deteksi *endpoint* terletak diawal sebelum proses ekstraksi fitur. Selain deteksi *endpoint*, praproses juga digunakan untuk meningkatkan sinyal suara. Algoritma deteksi *endpoint* menggunakan *short-time energy*. *Short-time energy*digambarkan dengan nilai/angka dari energi sinyal suara. Kuantitas sinyal suara didapatkan dari kuadrat amplitudo [10].

B. *Preemphasis, Frame Blocking, Windowing, dan FFT*

Proses *preemphasis, frame blocking*, *windowing*, dan FFT bertujuan untuk memperlakukan sinyal suara agar siap diubah menjadi informasi frekuensi yang berupa magnitude frekuensi. Preemphasis digunakan untuk mengingkatkan energi frekuensi tinggi dari sinyal suara. Cara kerja *preemphasis* hampir sama dengan *high-pass filter* yaitu untuk melewatkan komponen frekuensi tinggi dari sinyal suara [3].

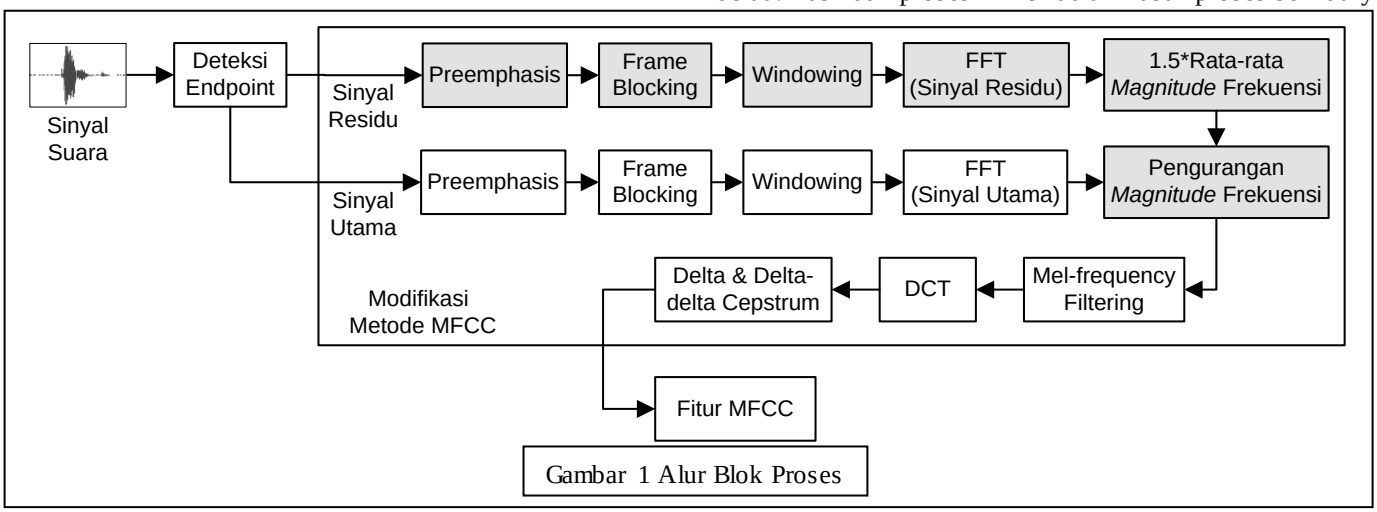
Frame blocking bertujuan untuk memecah sinyal suara menjadi blok *frame* kecil. *Frame* tadi berisi sampel yang cukup untuk mendapatkan informasi [2]. Sedangkan *windowing*berfungsi untuk mengurangikesenjangandi awal dan akhir dari *frame windowing*[6]. Ketika berada di *window*, sinyal suara pada awal dan akhir *frame* dibuat menjadi runcing [2]. Adapun *Fast Fourier Transform* (FFT) digunakan untuk merubah domain sinyal suara yang awalnya berasal dari domain waktu ke domain frekuensi [2]. Prose FFT menghasilkan magnitude frekuensi dalam bentuk bilangan kompleks. Bilangan kompleks tersebut terdiri dari riil dan imajiner kemudian dirubah menjadi bilangan riil dengan menghitung nilai mutlak magnitude.

C. *Rata-rata Magnitude Frekuensi*

Pada proses *frame blocking* dihasilkan beberapa blok sinyal suara. Setiap blok sinyal suara diubah menjadi magnitude frekuensi. Setiap frekuensi sinyal suara dari semua blok dirata-rata nilai magnitude-nya. Pada proses awal sinyal suara masuk kemudian dipisah antara sinyal utama dengan sinyal derau. Pada sinyal derau, nilai rata-rata tersebut digunakan sebagai estimasi nilai frekuensi derau.

D. *Pengurangan Magnitude Frekuensi*

Setiap *frame*/blok sinyal suara dihasilkan nilai magnitude frekuensi. Magnitude frekuensi sinyal utama dikurangi magnitude frekuensi sinyal derau [14][12]. Dari proses pengurangan dihasilkan magnitude frekuensi yang bersih dari derau. Hasil dari proses ini kemudian masuk proses berikutnya



Gambar 1 Alur Blok Proses

untuk dihasilkan koefisien fitur sinyal suara yang bersih dari derau.

III. HASIL DAN PEMBAHASAN

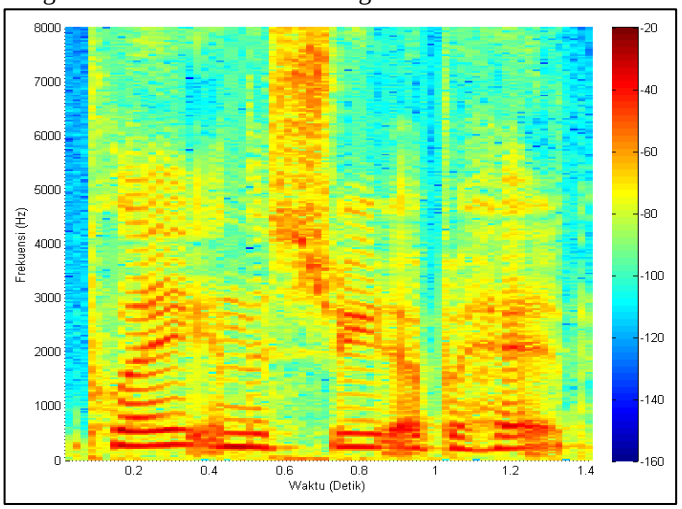
Pengujian dilakukan pada tipe derau dan tingkat SNR yang berbeda. Ada 5 tipe derau yang digunakan yaitu derau *f16* (pesawat tempur), *volvo* (kendaraan berat), *hfchannel* (frekuensi televisi), *white*, dan *pink*. Tingkat *Signal Noise Ratio* (SNR) yang digunakan sebesar 0, 5, 10, 15, 20, 25, dan bersih. Satu dataset terdiri dari 36 pembicara dan setiap pembicara terdiri dari 11 file. Pelatihan menggunakan 10 file pada setiap pembicara dari dataset sinyal yang bersih. Pengenalan menggunakan 1 file pada setiap pembicara dari dataset sinyal yang berderau. Pelatihan dan pengenalan untuk setiap file diulangi sebanyak 5 kali untuk mendapatkan akurasi maksimum. Pada uji coba dilakukan percobaan menggunakan metode yang diusulkan, MFCC, MFCC+SS, dan MFCC+wiener dengan jumlah komponen GMM sebesar 8, 16, 32, dan 64.

Nilai akurasi dari metode yang diusulkan mayoritas lebih tinggi dari metode lain kecuali pada SNR tingkat SNR 10dB. Pada SNR tingkat SNR 10dB, nilai akurasi tertinggi terletak pada metode MFCC+SS sebesar 99.6% dengan selisih 0.05% dari metode yang diusulkan. Selisih metode yang diusulkan dengan metode MFCC+SS sangat kecil sehingga akurasi kedua metode tersebut bisa dianggap sama. Dengan demikian,bisadisimpulkan bahwa metode yang diusulkan unggul pada semua tingkat SNR. Akurasi metode ekstraksi fitur terhadap tingkat SNR ditunjukkan pada tabel 1.

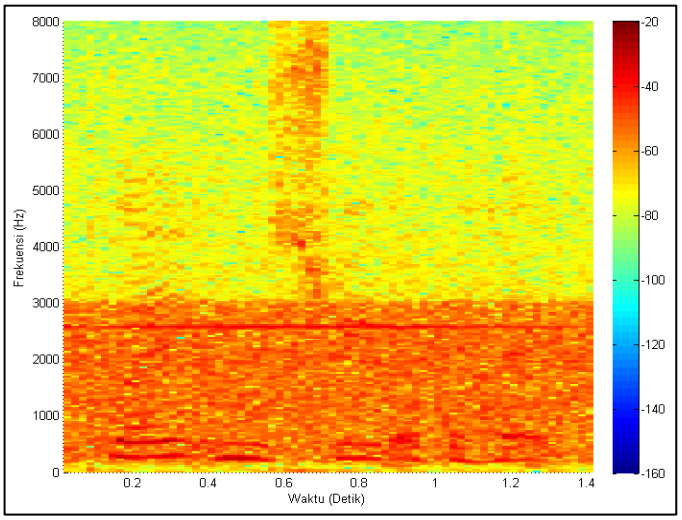
Tabel 1. Akurasi Metode Ekstraksi Fitur Terhadap Tingkat SNR (dalam satuan %)

Tingkat SNR	Metode yang diusulkan	MFCC	MFCC+ SS	MFCC+ wiener
Bersih	100	100	100	95.71
25dB	100	100	100	95.61
20dB	100	100	100	95.25
15dB	100	98.48	100	95.15
10dB	99.55	80.15	99.6	92.88
5dB	94.04	48.94	88.18	82.88
0dB	66.46	29.6	53.08	57.78
Ket: Nilai akurasi tertinggi				

Penurunan akurasi seiring turunnya SNR tingkat SNR disebabkan karena frekuensi pembicara tertutup oleh frekuensi derau. Frekuensi derau dengan magnitude yang lebih besar menutupi frekuensi pembicara yang mempunyai magnitude lebih kecil.Sebagai ilustrasi dapat dilihat gambar spektrogram dari pembicara 1 pada Gambar 2 dan Gambar 3. Pada Gambar 2 dan Gambar 3 terdapat gambar spektrogram dari pembicara 1 pada derau hfchannel dengan tingkat SNR bersih dan 0dB.Ketika tingkat SNR bersih, frekuensi pembicara terlihat jelas. Kemudian frekuensi pembicara tersamarkan dan mirip dengan frekuensi derau ketika tingkat SNR 0dB.



Gambar 2. Spektrogram Pembicara 1 tanpa Derau



Gambar 3. Spektrogram Pembicara 1 pada Derau Hfchannel dengan SNR 0dB

Dari hasil analisis yang sudah dilakukan, metode yang diusulkan unggul pada semua analisis. Metode yang diusulkan unggul pada 23 pembicara dari 36 pembicara. Metode yang diusulkan menghasilkan nilai akurasi lebih tinggi pada beberapa tipe derau kecuali tipe derau volvo yang sama dengan metode MFCC dan MFCC+SS. Pada kondisi tingkat SNR bermacam-macam, metode yang diusulkan unggul pada semua tingkat SNR kecuali pada SNR tingkat SNR 10dB yang lebih unggul metode MFCC+SS dengan selisih 0.05% dari metode yang diusulkan. Rata-rata akurasi metode yang diusulkan, MFCC, MFCC+SS, dan MFCC+wiener secara berturut-turut adalah 94.29%, 79.6%, 91.55%, dan 87.89%. Dengan demikian dapat disimpulkan bahwa metode yang diusulkan memiliki kinerja lebih baik dari metode MFCC, MFCC+SS, dan MFCC+wiener dalam hal akurasi.

Metode yang diusulkan memiliki kinerja yang lebih baik dari metode MFCC, MFCC+SS, dan MFCC+wiener karena metode yang diusulkan menggunakan residu (sisa hasil) algoritma deteksi *endpoint* sebagai informasi frekuensi derau. Residu deteksi *endpoint* diolah untuk mendapatkan informasi frekuensi derau. Sinyal utama dikurangi dengan frekuensi derau sehingga sinyal utama tidak mengandung derau lagi.

IV. KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari penelitian ini adalah modifikasi metode MFCC menggunakan sinyal residu dari algoritma deteksi endpoint dapat meningkatkan kinerja sistem dalam hal akurasi. Metode yang diusulkan, MFCC, MFCC+SS, dan MFCC+wiener berturut-turut memiliki rata-rata akurasi 94.29%, 79.6%, 91.55%, dan 87.89%. Metode yang diusulkan menunjukkan peningkatan akurasi berturut-turut sebesar 14.69%, 2.74%, dan 6.4% dari metode MFCC, MFCC+SS, dan MFCC+wiener. Akurasi dari metode ekstraksi fitur turun seiring turunnya tingkat SNR. Metode yang diusulkan unggul hampir pada semua tingkat SNR kecuali pada tingkat SNR 10dB. Pada tingkat SNR 10dB, metode MFCC+SS memiliki akurasi tertinggi sebesar 99.6% dengan selisih 0.05% dari metode yang diusulkan

DAFTAR PUSTAKA

[1] P. M. Chauhan and N. P. Desai, “Mel Frequency Cepstral Coefficients (MFCC) based speaker identification in noisy environment using wiener filter.”

[2] S. Gupta, J. Jaafar, W. wan Ahmad, and A. Bansal, “Feature Extraction Using Mfcc.”

[3] Wei Han, Cheong-Fat Chan, Chiu-Sing Choy, and Kong-Pang Pun, “An efficient MFCC extraction method in speech recognition.”

[4] J. Wu and J. Yu, “An improved arithmetic of MFCC in speech recognition system.”

[5] T. Kinnunen *et al.*, “Low-variance multitaper MFCC features: A case study in robust speaker verification.”

[6] C. G. K. Leon, “Robust computer voice recognition using improved MFCC algorithm.”

[7] J. S. Lim and A. V. Oppenheim, “Enhancement and Bandwidth Compression of Noisy Speech.”

[8] M. Jaafar, A. Ahmad, Z. Sakawi, M. Abdullah, N. Sulaiman, and Normukhnun Mokhtar, “Indeks Kualiti Air (IKA) Sg . Selangor pasca pembinaan Water Quality Index (WQI) of the Selangor River after the construction of the Selangor River Dam.”

[9] Bing-Fei Wu and Kun-Ching Wang, “Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments.”

[10] G. Rigoll, “Speaker adaptation using improved speaker Markov models.”

[11] C. Leimin, “Performance Comparison of New Endpoint Detection Method in Noise Environments.”

[12] Y. Wang, B. Li, X. Jiang, F. Liu, and L. Wang, “Speaker recognition based on dynamic MFCC parameters.”

[13] Y. Zhang and W. H. Abdulla, “Robust speaker identification in noisy environment using cross diagonal GTF-ICA feature.”

[14] W. U. Zunjing and C. A. O. Zhigang, “Improved MFCC-based feature for robust speaker identification.”