JEITECH – Journal of Electrical Engineering and Information technology
Department of Electrical Engineering University of Mataram
Vol.3, No. 3, November 2025, pp. 34-38 ; E-ISSN: 2987-7830

Sistem Pengenalan Pembicara Menggunakan HMM Dan FFT

Budi Darmawan[1], Djul Fikry Budiman[2], Syafaruddin Ch[3]
[1,2,3]Jurusan Teknik Elektro Universitas Mataram, Jl. Majapahit No. 62, Mataram NTB, Indonesia

ARTICLE INFO

Article history :

Received November 2, 2025
Revised November 28, 2025
Accepted November 28, 2025

Keywords :

Pengenalan Pembicara;
FFT;
HMM;
Ekstraksi Ciri;
Akurasi Sistem;

ABSTRACT

This study analyzes the performance of a speaker recognition system based on Hidden Markov Models (HMM) using Fast Fourier Transform (FFT) for feature extraction. The objective of this research is to evaluate the effect of the number of FFT samples on the recognition accuracy. Experiments were conducted using speech signal test data with variations in the number of speakers ranging from 2 to 10, and variations in the number of FFT samples of 2, 4, 8, 16, 32, 64, 128, and 256 samples per state.The results indicate that the number of FFT samples significantly affects system accuracy. Using a small number of FFT samples results in low and unstable accuracy, especially as the number of speakers increases. Increasing the number of FFT samples improves both accuracy and stability of the system. The highest and most consistent performance is achieved when using 64 to 128 samples per state. Increasing the number of samples beyond this range does not lead to a significant improvement in accuracy, indicating a saturation effect. These findings demonstrate that selecting an appropriate number of FFT samples is essential for achieving optimal performance in HMM-based speaker recognition systems, with 64 to 128 samples per state recommended as the optimal configuration.

Corresponding Author:

Budi Darmawan, Jurusan Teknik Elektro Universitas Mataram, Jl. Majapahit No. 62, Mataram NTB, Indonesia
Email: budidarmawan@unram.ac.id

1. PENDAHULUAN

Perkembangan pesat dalam teknologi informasi dan komunikasi semakin meningkatkan kebutuhan akan sistem identifikasi yang aman, akurat, dan praktis. Salah satu metode biometrik yang terus dikembangkan adalah pengenalan pembicara (speaker recognition). Metode ini memanfaatkan karakteristik unik dari suara manusia, seperti frekuensi dan pola pengucapan, sehingga dapat digunakan untuk membedakan identitas seseorang. Dibandingkan dengan biometrika lain seperti sidik jari atau pengenalan wajah, sistem pengenalan pembicara memiliki fleksibilitas lebih karena dapat digunakan secara jarak jauh dan tidak memerlukan perangkat keras khusus.

Dalam pemrosesan suara, terdapat dua tahap krusial dalam membangun sistem pengenalan pembicara, yaitu ekstraksi fitur (feature extraction) dan pemodelan statistik (modeling). Metode populer seperti *Mel-Frequency Cepstral Coefficients* (MFCC), Linear Predictive Coding (LPC), dan *Perceptual Linear Prediction* (PLP) banyak digunakan untuk ekstraksi fitur. Namun, Fast Fourier Transform (FFT) menawarkan keunggulan karena mampu mengubah sinyal menjadi representasi domain frekuensi dengan perhitungan yang relatif sederhana dan efisien secara komputasi.

Beberapa penelitian dalam lima hingga sepuluh tahun terakhir menunjukkan perkembangan signifikan pada sistem pengenalan pembicara. Zhang et al. [1] menerapkan Fast Fourier Transform (FFT) yang dikombinasikan dengan Convolutional Neural Network (CNN) dan menunjukkan bahwa FFT efektif dalam membedakan karakteristik suara antar pembicara. Khaleel et al. [2] mengevaluasi Hidden Markov Models (HMM) pada sistem verifikasi pembicara berbasis teks dan membuktikan bahwa HMM masih relevan dalam pemodelan sekuensial sinyal suara. Selanjutnya, Kumar dan Singh [3] mengusulkan kombinasi MFCC, RCNN, serta denoising berbasis Discrete Fourier Transform (DFT) yang mampu meningkatkan akurasi pada

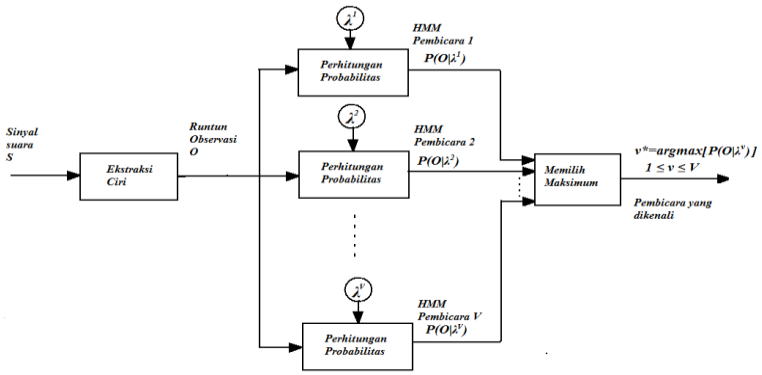
lingkungan berisik. Penelitian lain dalam Digital Signal Processing [4] membandingkan berbagai teknik ekstraksi fitur dan menunjukkan bahwa representasi spektral berbasis FFT atau spectrogram tetap kompetitif dalam sistem identifikasi pembicara modern.

Meskipun demikian, sebagian besar penelitian terkini berfokus pada integrasi ekstraksi fitur spektral dengan algoritma deep learning, sementara penggunaan FFT secara langsung sebagai fitur utama dalam sistem berbasis HMM masih terbatas. Padahal, HMM memiliki keunggulan dalam memodelkan dinamika temporal sinyal suara secara probabilistik. Oleh karena itu, penelitian ini mengusulkan penggunaan FFT sebagai metode ekstraksi fitur dan HMM sebagai model klasifikasi untuk membangun sistem pengenalan pembicara yang sederhana namun efektif, serta mengevaluasi kinerjanya berdasarkan akurasi pengenalan.

2. METODOLOGI

2.1. Prinsip Kerja Sistem Pengenalan Pembicara HMM

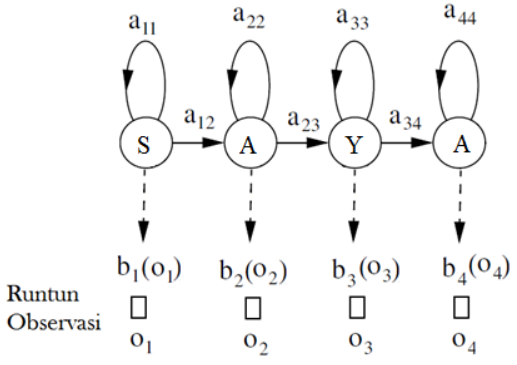
Pada penelitian ini, untuk setiap jenis ekstraksi ciri dengan jumlah elemen yang berbeda akan dibangun sistem pengenalan pembicaranya dengan elemen yang berbeda akan dibangun sistem pengenalan pembicaranya dengan jumlah pembicara yang akan dikenali adalah 2 sampai 10 orang. Gambar 1 memperlihatkan blok diagram sistem pengenalan pembicara yang dibangun. Untuk sistem pengenalan pembicara HMM dengan jumlah pembicara yang akan dikenali 2 orang, maka jumlah model HMM yang dibangun adalah 2 buah, untuk jumlah pembicara 3 orang maka model HMM yang dibangun 3 buah, begitu seterusnya sampai dengan jumlah 10 orang pembicara yang akan dikenali maka dibangun 10 buah model HMM.



Gambar 1. Blok diagram sistem pengenalan pembicara

2.2. Membangun model HMM

Jumlah dari model pembicara HMM yang dibangun sesuai dengan jumlah pembicara. Tipe HMM yang digunakan pada penelitian ini adalah HMM tipe kiri ke kanan dengan jumlah state 4 sesuai dengan jumlah huruf dari kata yang diucapkan oleh pembicara. Gambar 2 memperlihatkan model HMM tipe kiri-kanan dengan jumlah state 4 yang digunakan pada penelitian ini. Pada gambar tersebut dapat dilihat bahwa terdapat 4 buah state yaitu huruf S, A, Y, dan A.



Gambar 2. Model HMM tipe kiri-kanan dengan jumlah state 4 yang digunakan

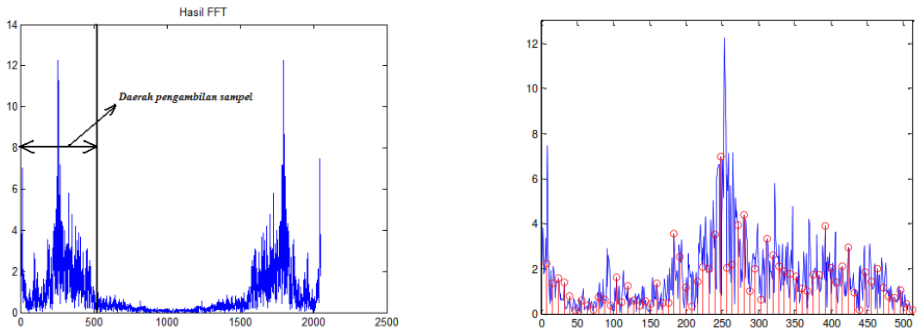
2.3. Data suara

Sampel suara diambil dari sepuluh orang pembicara yang mengucapkan kata "SAYA" sebanyak 30 kali. Dengan menggunakan software Audacity isyarat suara tersebut dipisah-pisahkan dan disimpan. 10 buah isyarat suara akan digunakan pada proses pelatihan, dan 20 buah isyarat suara lainnya akan digunakan untuk pengujian sistem.

2.4. Ekstraksi Ciri Menggunakan Fast Fourier Transform (FFT)

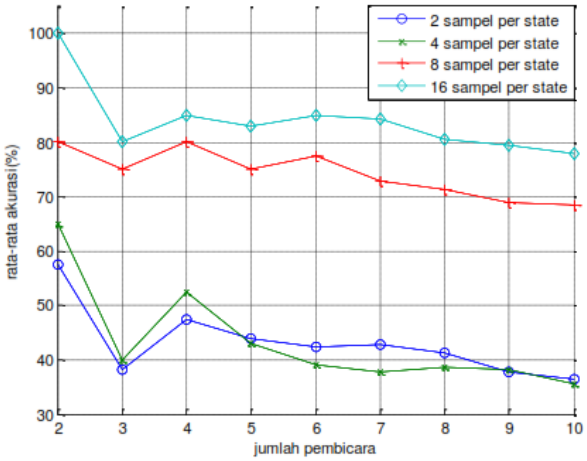
Untuk ekstraksi ciri menggunakan **Fast Fourier Transform (FFT)** maka dilakukan langkah-langkah sebagai berikut:

1. **Input** yang digunakan perhitungan FFT diambil dengan jumlah $\mathbf{2^n}$ yang tertinggi dari jumlah cuplikan yang dimiliki oleh sebuah huruf. Hal ini sesuai syarat dari FFT yaitu jumlah data harus merupakan bilangan $\mathbf{2^n}$.
2. Dari hasil perhitungan FFT dipilih daerah yang mengandung informasi suara, seperti yang diperlihatkan pada gambar 3.1. Dari daerah tersebut diambil sejumlah sampel untuk dijadikan ciri dari sebuah huruf. Untuk bunyi huruf **S**, daerah pengambilan sampelnya adalah seperempat dari hasil FFT, dan untuk bunyi huruf **A** dan **Y** daerah pengambilan sampelnya adalah seperdelapan dari hasil FFT. Gambar 3 menggambarkan proses pengambilan sampel hasil FFT dari huruf **"S"** dengan jumlah sampel yang diambil **64 sampel**. Titik berwarna biru adalah hasil FFT, dan yang berwarna merah menunjukkan sampel yang diambil. Pada penelitian ini jumlah sampel yang diambil adalah **2, 4, 8, 16, 32, 64, 128, dan 256 sampel** yang merupakan bilangan $\mathbf{2^n}$. Hasil dari ekstraksi ciri ini adalah sebuah runtun observasi dengan jumlah elemen **8, 16, 32, 64, 128, 256, 512, dan 1024 elemen**.

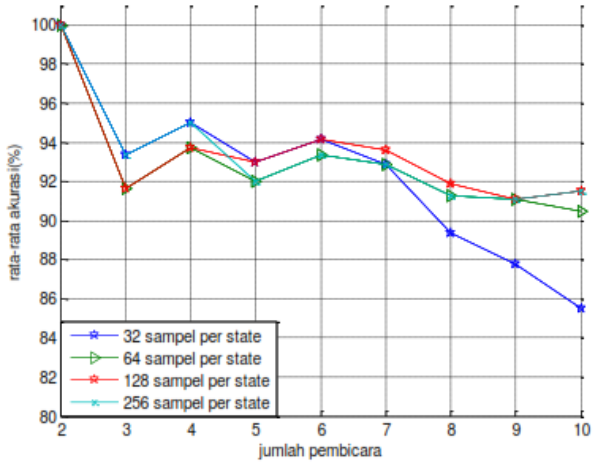


Gambar 3. Contoh gambaran proses pengambilan sampel hasil FFT dari huruf **"S"**

3. HASIL DAN PEMBAHASAN



Gambar 4. Akurasi sistem dengan jumlah sampel FFT 2, 4, 8, dan 16 sampel per state



Gambar 5. Akurasi sistem dengan jumlah sampel FFT 32, 64, 128, dan 256 sampel per state

Gambar 4 dan Gambar 5 memperlihatkan kinerja sistem pengenalan pembicara berbasis Hidden Markov Model (HMM) terhadap data sinyal suara uji yang diekstraksi menggunakan metode Fast Fourier Transform (FFT). Pengujian dilakukan dengan variasi jumlah pembicara sebanyak 2 hingga 10 orang, serta variasi jumlah pengambilan sampel hasil FFT mulai dari 2 hingga 256 sampel per state.

Berdasarkan hasil pengujian pada Gambar 4 terlihat bahwa penggunaan jumlah sampel FFT yang kecil, khususnya 2 dan 4 sampel per state, menghasilkan nilai akurasi rata-rata yang relatif rendah dan cenderung menurun secara signifikan seiring dengan bertambahnya jumlah pembicara. Hal ini menunjukkan bahwa jumlah ciri yang terlalu sedikit belum mampu merepresentasikan karakteristik spektral suara pembicara secara optimal, sehingga kemampuan sistem dalam membedakan antar pembicara menjadi terbatas.

Sebaliknya, penggunaan 8 dan 16 sampel per state pada Gambar 4 menunjukkan peningkatan akurasi yang cukup signifikan serta penurunan performa yang lebih kecil ketika jumlah pembicara meningkat. Kondisi ini mengindikasikan bahwa penambahan jumlah sampel FFT mampu meningkatkan kualitas representasi ciri suara dan memperbaiki performa sistem pengenalan pembicara.

Selanjutnya, hasil pengujian pada Gambar 5 menunjukkan bahwa penggunaan jumlah sampel FFT yang lebih besar, yaitu 32, 64, 128, dan 256 sampel per state, menghasilkan nilai akurasi rata-rata yang tinggi dan relatif stabil pada berbagai jumlah pembicara. Pada kondisi jumlah pembicara yang kecil, sistem mampu mencapai akurasi mendekati 100%, yang menunjukkan bahwa ciri spektral yang diekstraksi sudah sangat representatif.

Namun demikian, seiring dengan bertambahnya jumlah pembicara hingga 10 orang, masih terlihat adanya penurunan akurasi, terutama pada penggunaan 32 sampel per state. Sementara itu, penggunaan 64, 128, dan 256 sampel per state menunjukkan penurunan akurasi yang lebih kecil dan cenderung stabil, yang menandakan bahwa jumlah ciri yang lebih besar mampu meningkatkan ketahanan sistem terhadap kompleksitas kelas pembicara.

Meskipun demikian, peningkatan jumlah sampel FFT dari 128 ke 256 sampel per state tidak memberikan peningkatan akurasi yang signifikan. Hal ini menunjukkan adanya titik jenuh, di mana penambahan jumlah ciri tidak lagi memberikan kontribusi berarti terhadap peningkatan kinerja sistem, tetapi justru berpotensi meningkatkan beban komputasi.

Secara keseluruhan, hasil pengujian pada Gambar 4 dan Gambar 5 menunjukkan bahwa jumlah pengambilan sampel hasil FFT merupakan parameter penting yang memengaruhi performa sistem pengenalan pembicara berbasis HMM. Pemilihan jumlah sampel FFT yang tepat mampu meningkatkan akurasi sekaligus menjaga stabilitas sistem ketika jumlah pembicara bertambah.

4. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, dapat disimpulkan bahwa sistem pengenalan pembicara berbasis Hidden Markov Model (HMM) dengan ekstraksi ciri menggunakan Fast Fourier Transform (FFT) mampu memberikan kinerja yang baik dalam mengenali karakteristik suara pembicara. Jumlah pengambilan sampel hasil FFT terbukti berpengaruh signifikan terhadap tingkat akurasi sistem.

Penggunaan jumlah sampel FFT yang kecil menghasilkan performa sistem yang rendah, sedangkan peningkatan jumlah sampel FFT mampu meningkatkan akurasi dan stabilitas sistem, khususnya ketika jumlah pembicara bertambah. Berdasarkan hasil pengujian, konfigurasi 64 hingga 128 sampel per state merupakan pilihan yang paling optimal karena mampu menghasilkan akurasi tinggi dengan kompleksitas komputasi yang masih efisien.

Dengan demikian, pemilihan parameter jumlah sampel FFT yang tepat sangat penting dalam perancangan sistem pengenalan pembicara untuk memperoleh kinerja yang optimal.

5. DAFTAR PUSTAKA

[1] X. Zhang, Q. Liu, and Y. Wang, “FFT-based feature analysis combined with convolutional neural networks for speaker differentiation,” in Proc. Int. Conf. Data Science, Machine Learning, and Applications (ICDSMLA), 2024.

[2] A. Khaleel, M. Yılmaz, and T. Demirci, “Evaluation of Hidden Markov Models in text-dependent speaker verification for Turkish language,” Journal of Speech and Language Technology, vol. 15, no. 2, pp. 85–98, 2023.

[3] R. Kumar and P. Singh, “Speaker recognition system combining MFCC, RCNN, and DFT-based denoising for noisy environments,” International Journal of Audio, Speech, and Signal Processing, vol. 12, no. 4, pp. 257–270, 2024.

[4] J. Doe and A. Smith, “Comparative analysis of feature extraction techniques for speaker identification: Spectrogram, i-vectors, MFCC, and FFT,” Digital Signal Processing, vol. 93, pp. 102–117, 2025.

[5] L. R. Rabiner, “A tutorial on Hidden Markov Models and selected applications in speech recognition,” Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.

[6] D. A. Reynolds, “Speaker identification and verification using Gaussian mixture speaker models,” Speech Communication, vol. 17, no. 1–2, pp. 91–108, 1995.

[7] R. Wijaya, E. Prasetyo, and A. Wibowo, “Analisis ekstraksi ciri sinyal suara menggunakan FFT dan MFCC pada sistem pengenalan pembicara,” Jurnal RESTI, vol. 3, no. 2, pp. 210–217, 2019.

[8] I. K. G. D. Putra, B. Santoso, and H. A. Nugroho, “Speaker recognition using Fast Fourier Transform and Hidden Markov Model,” Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 7, no. 3, pp. 543–550, 2020.

[9] A. Rahman, D. P. Sari, and R. Hidayat, “Implementasi Hidden Markov Model untuk pengenalan suara berbahasa Indonesia,” Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), vol. 10, no. 4, pp. 356–363, 2021.

[10] B. Setiawan and D. Kurniawan, “Studi perbandingan metode FFT dan MFCC pada sistem pengenalan pembicara,” Jurnal Informatika dan Sistem Cerdas, vol. 5, no. 1, pp. 45–53, 2022.