



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 2 (2026) pp: 16881-16889

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Ekstraksi Ciri Suara Gender Berdasarkan Frequency Domain Audio Feature

Dhiandra Nirina Yusuf, Hari Purwadi, Didi Susilo Budi Utomo

Program Studi Teknik Komputer, Politeknik Negeri Samarinda

dhiandranirinaa@gmail.com, hari.purwa06@gmail.com

Abstrak

Penelitian ini bertujuan mengekstraksi ciri suara berdasarkan gender menggunakan Mel-Frequency Cepstral Coefficients (MFCC) dan Spectral Centroid. Data yang digunakan adalah 20 sampel rekaman vokal "A" dari 4 responden (2 laki-laki, 2 perempuan), masing-masing 5 kali perekaman. Seluruh sampel di-resample ke 16.000 Hz. Ekstraksi MFCC menghasilkan 13 koefisien yang dirata-ratakan per sampel, sedangkan Spectral Centroid dihitung rata-rata dan standar deviasinya. Hasil pengujian statistik dengan Independent Samples t-Test pada taraf signifikansi 0,05 menunjukkan perbedaan yang signifikan secara statistik antara kelompok laki-laki dan perempuan pada koefisien MFCC-2 ($p=0,0154$), MFCC-5 ($p=0,0080$), MFCC-7 ($p=0,0014$), MFCC-12 ($p<0,0001$), dan MFCC-13 ($p=0,0001$). Sementara itu, MFCC-1 tidak menunjukkan perbedaan yang signifikan ($p=0,0723$). Hasil perhitungan Spectral Centroid menunjukkan bahwa rata-rata perempuan mencapai 3105,38 Hz dan laki-laki 1944,91 Hz, dengan standar deviasi yang lebih tinggi pada perempuan (1113,47 Hz vs 486,80 Hz). Temuan ini mengonfirmasi bahwa suara perempuan memiliki distribusi energi frekuensi yang cenderung lebih tinggi dan lebih bervariasi, sesuai dengan karakteristik fisiologis pita suara yang lebih pendek. Informasi spektral ini memperlihatkan potensi MFCC dan Spectral Centroid sebagai penciri akustik gender yang dapat diaplikasikan pada asisten virtual atau forensik audio. Kedua fitur mampu merepresentasikan perbedaan akustik gender, meskipun penelitian terbatas pada sampel kecil. Oleh karena itu, diperlukan penelitian lanjutan dengan dataset lebih besar, variasi fonem, rentang usia, dan lingkungan perekaman terkontrol untuk meningkatkan keandalan fitur.

Kata kunci: Ekstraksi Ciri Suara, MFCC, Spectral Centroid, Pengolahan Sinyal Suara, Gender

1. Latar Belakang

Perkembangan teknologi informasi dan kecerdasan buatan telah mendorong pemanfaatan sistem pengenalan suara dalam berbagai bidang, seperti asisten virtual, sistem keamanan, dan interaksi manusia-komputer. Agar sistem tersebut dapat mengenali dan memproses suara secara efektif, diperlukan teknik pengolahan sinyal yang mampu mempresentasikan karakteristik penting dari suara yang diterima. Namun, sinyal suara yang diperoleh secara langsung umumnya memiliki ukuran data yang besar dan mengandung berbagai informasi yang tidak semuanya relevan untuk proses pengenalan [1]. Oleh karena itu, diperlukan suatu proses ekstraksi ciri (feature extraction) untuk mengambil karakteristik utama dari sinyal suara sehingga informasi yang diperoleh menjadi lebih representatif dan mudah diolah oleh sistem [2].

Ekstraksi ciri merupakan tahap penting dalam proses analisis audio karena berfungsi mengubah sinyal suara mentah menjadi sekumpulan fitur yang mampu merepresentasikan identitas atau karakteristik suara secara efektif. Secara umum, proses ekstraksi ciri adalah langkah penting dalam pengenalan pola dan berbagai tugas machine learning. Tujuannya adalah untuk menyaring atau mengekstrak serangkaian fitur dari dataset yang sedang diamati atau diteliti. Fitur yang dihasilkan melalui proses ekstraksi ciri harus mampu mempertahankan informasi yang penting serta merepresentasikan karakteristik utama dari data asli secara efektif [3].

Suara merupakan gelombang mekanik yang dihasilkan oleh getaran suatu objek dan merambat melalui media seperti udara, cairan, maupun benda padat [4]. Pada manusia, suara dihasilkan oleh getaran pita suara yang kemudian merambat melalui udara dan dapat diterima oleh sistem pendengaran manusia. Suara yang dihasilkan mengandung berbagai informasi, seperti identitas pembicara, emosi, serta karakteristik fisiologis tertentu yang dapat dianalisis melalui teknik pengolahan sinyal digital [5].

Dalam pengolahan sinyal digital, suara direpresentasikan dalam bentuk sinyal audio yang dapat dibedakan menjadi sinyal analog dan sinyal digital. Sinyal analog merupakan sinyal yang bersifat kontinu terhadap waktu dan memiliki rentang nilai amplitudo yang tidak terbatas. Sementara itu, sinyal digital merupakan representasi diskrit dari sinyal analog yang diperoleh melalui proses digitalisasi sehingga memiliki nilai amplitudo yang terbatas [6]. Umumnya, suara yang dapat didengar oleh manusia berada pada rentang frekuensi 20 Hz hingga 20.000 Hz. Representasi digital dari sinyal suara memungkinkan berbagai proses analisis dilakukan secara lebih efisien, termasuk ekstraksi ciri pada domain frekuensi untuk mengidentifikasi karakteristik tertentu dari suatu suara.

Berbagai metode ekstraksi ciri telah dikembangkan, diantaranya *Mel-Frequency Cepstral Coefficients (MFCC)* dan *Spectral Centroid*. *MFCC* banyak digunakan karena mampu merepresentasikan karakteristik spektral suara yang sesuai dengan persepsi pendengaran manusia [7], sedangkan *Spectral Centroid* digunakan untuk menggambarkan pusat distribusi energi pada spektrum frekuensi suatu sinyal audio [8]. Kedua fitur tersebut dapat digunakan untuk mengidentifikasi perbedaan karakteristik suara berdasarkan gender.

Mel-Frequency Cepstral Coefficients (MFCC) merupakan salah satu metode ekstraksi ciri yang banyak digunakan dalam pengolahan sinyal suara dan sistem pengenalan suara. Metode ini dirancang berdasarkan karakteristik sistem pendengaran manusia dengan menggunakan skala Mel, yaitu skala frekuensi yang merepresentasikan cara manusia memersepsikan perbedaan frekuensi suara [9]. Oleh karena itu, *MFCC* mampu menghasilkan representasi spektral yang lebih sesuai dengan persepsi pendengaran manusia dibandingkan representasi frekuensi linear.

Proses ekstraksi ciri menggunakan *MFCC* dilakukan melalui beberapa tahapan, yaitu *pre-emphasis*, *framing*, *windowing*, transformasi ke domain frekuensi menggunakan Fast Fourier Transform (FFT), pembentukan Mel Filter Bank, dan Discrete Cosine Transform (DCT). Hasil dari proses tersebut berupa sekumpulan koefisien yang dapat merepresentasikan karakteristik penting dari sinyal suara secara efektif [10]. Karena kemampuannya dalam menangkap informasi spektral suara, *MFCC* banyak digunakan dalam berbagai penelitian yang berkaitan dengan pengenalan pembicara, pengenalan ucapan, serta analisis karakteristik suara [7].

Spectral Centroid merupakan salah satu fitur dalam domain frekuensi yang digunakan untuk menggambarkan pusat distribusi energi pada spektrum suatu sinyal audio. Fitur ini menunjukkan lokasi rata-rata energi frekuensi yang terdapat dalam sinyal sehingga sering dikaitkan dengan persepsi kecerahan (*brightness*) suatu suara [11]. Semakin tinggi nilai *Spectral Centroid*, maka persepsi pendengaran kita akan menangkap suara tersebut sebagai suara yang cerah “bright” atau tajam. Sebaliknya, nilai yang rendah merujuk pada suara yang gelap “dark” atau berat. Oleh karena itu, fitur ini sangat efektif untuk membedakan warna suara (timbre) dalam analisis audio [3].

Nilai *Spectral Centroid* dapat dihitung menggunakan persamaan berikut:

$$SC = \frac{\sum_{k=1}^N f_k X(k)}{\sum_{k=1}^N X(k)}$$

Keterangan:

SC = nilai *Spectral Centroid*,

f_k = frekuensi pada bin ke- k ,

$X(k)$ = magnitudo spektrum pada frekuensi ke- k ,

N = jumlah bin frekuensi.

Dalam penelitian pengolahan sinyal suara, *Spectral Centroid* banyak digunakan untuk menganalisis karakteristik spektral karena mampu memberikan informasi mengenai distribusi energi frekuensi pada suatu sinyal audio [12]. Oleh karena itu, fitur ini dapat dimanfaatkan untuk mengamati perbedaan karakteristik suara berdasarkan gender melalui kecenderungan distribusi frekuensi yang dimiliki oleh masing-masing kelompok suara.

Perbedaan karakteristik suara antara laki-laki dan perempuan dipengaruhi oleh faktor fisiologis pada sistem produksi suara, seperti ukuran laring, panjang dan ketebalan pita suara, serta bentuk saluran vokal. Secara umum, suara laki-laki memiliki pita suara yang lebih panjang dan tebal sehingga menghasilkan frekuensi dasar (fundamental frequency) yang lebih rendah dibandingkan suara perempuan. Sebaliknya, suara perempuan cenderung memiliki frekuensi dasar yang lebih tinggi karena pita suara yang lebih pendek dan tipis. Perbedaan

fisiologis tersebut menyebabkan distribusi energi dan karakteristik spektral yang berbeda pada sinyal suara, sehingga dapat diamati melalui berbagai fitur yang diekstraksi dari domain frekuensi [13].

Analisis karakteristik suara berdasarkan gender menjadi salah satu topik yang menarik dalam bidang pengolahan sinyal suara karena dapat memberikan pemahaman mengenai perbedaan pola akustik yang dimiliki oleh masing-masing kelompok suara. Informasi tersebut juga dapat dimanfaatkan sebagai dasar dalam pengembangan berbagai sistem pengenalan suara dan penelitian terkait analisis audio.

Beberapa penelitian terdahulu telah memanfaatkan metode *Mel-Frequency Cepstral Coefficients (MFCC)* untuk mengekstraksi karakteristik suara pada berbagai aplikasi pengolahan sinyal suara. Salah satunya penelitian yang dilakukan oleh [9], menggunakan *MFCC* untuk menganalisis pola suara dan menunjukkan bahwa metode tersebut mampu mengubah sinyal suara menjadi vektor fitur yang dapat merepresentasikan karakteristik suara secara efektif. Selain itu, penelitian yg dilakukan [14], menunjukkan bahwa *MFCC* merupakan salah satu metode ekstraksi ciri yang banyak digunakan karena dirancang berdasarkan skala Mel yang sesuai dengan persepsi pendengaran manusia.

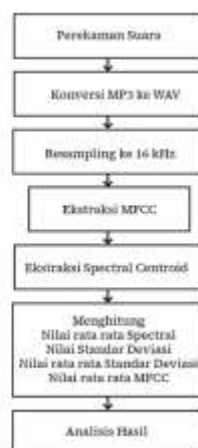
Berdasarkan uraian tersebut, penelitian ini bertujuan untuk melakukan ekstraksi ciri suara berdasarkan gender menggunakan fitur *Mel-Frequency Cepstral Coefficients (MFCC)* dan *Spectral Centroid*. Analisis dilakukan terhadap data suara laki-laki dan perempuan untuk mengamati karakteristik yang dihasilkan oleh masing-masing fitur. Hasil penelitian diharapkan dapat memberikan gambaran mengenai perbedaan karakteristik suara berdasarkan gender serta menunjukkan kemampuan kedua fitur tersebut dalam merepresentasikan informasi penting pada sinyal suara.

2. Metode Penelitian

Penelitian ini menggunakan perangkat keras berupa laptop dan telepon genggam sebagai alat perekaman suara. Adapun perangkat lunak yang digunakan meliputi Google Colaboratory sebagai lingkungan pemrograman, Python sebagai bahasa pemrograman utama, serta pustaka Librosa, NumPy, Pandas, dan Matplotlib untuk proses pengolahan data, ekstraksi ciri, dan visualisasi hasil penelitian.

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui proses perekaman suara secara langsung menggunakan telepon genggam. Data terdiri atas 20 sampel suara yang terbagi menjadi 10 sampel suara laki-laki dan 10 sampel suara perempuan. Sampel suara diperoleh dari 4 orang responden yang terdiri atas 2 laki-laki dan 2 perempuan, di mana setiap responden merekam pengucapan huruf vokal “A” sebanyak 5 kali.

Hasil perekaman awal disimpan dalam format MP3 dengan durasi rata-rata sekitar 1,17 detik. Selanjutnya seluruh data audio dikonversi ke format WAV menggunakan perangkat lunak VLC Media Player. Untuk menyeragamkan karakteristik data, seluruh sampel audio kemudian dilakukan proses resampling menjadi 16.000 Hz menggunakan Audacity sebelum digunakan pada tahap ekstraksi ciri. Tahapan penelitian yang dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Sistem

2.1. Perekaman Suara

Tahap perekaman suara dilakukan untuk memperoleh data audio yang siap digunakan pada proses ekstraksi ciri. Data penelitian berupa rekaman suara pengucapan huruf “A” yang diperoleh dari 2 responden laki laki dan 2 responden Perempuan. Setiap responden melakukan perekaman sebanyak 5 kali sehingga diperoleh total 20 sampel suara yang terdiri atas 10 sampel suara laki laki dan 10 sampel Perempuan. Hasil perekaman awal disimpan dalam format MP3 dengan durasi rata-rata sekitar 1,17 detik. Selanjutnya seluruh data dikonversi ke format WAV menggunakan VLC Media Player dan dilakukan proses resampling menjadi 16.000 Hz menggunakan Audacity. Proses ini bertujuan untuk menyeragamkan format dan frekuensi sampling seluruh data audio sehingga dapat diproses secara konsisten pada tahap ekstraksi ciri. Penyeragaman nilai sample rate dilakukan karena perbedaan keragaman pengambilan sampel dapat memengaruhi representasi frekuensi pada sinyal audio. Oleh karena itu, seluruh sampel suara disesuaikan menjadi 16.000 Hz agar karakteristik yang diperoleh dari proses ekstraksi ciri dapat dibandingkan secara lebih konsisten.

2.2. Ekstraksi MFCC

Proses ekstraksi ciri menggunakan metode *Mel-Frequency Cepstral Coefficients (MFCC)* dilakukan terhadap seluruh sampel suara yang telah melalui tahap persiapan data. Ekstraksi dilakukan menggunakan Pustaka Librosa dan Scipy pada bahasa pemrograman python yang dijalankan melalui google colaboratory. Tahapan ekstraksi diawali dengan proses framing dan pemberian Hamming Window pada setiap frame sinyal suara. Selanjutnya dilakukan transformasi Fast Fourier Transform (FFT) untuk memperoleh representasi spektrum frekuensi. Spektrum frekuensi kemudian diproses menggunakan Mel Filter Bank dan dikonversi ke skala logaritmik sebelum dilakukan Discrete Cosine Transform (DCT) untuk menghasilkan koefisien *MFCC*.

Hasil ekstraksi berupa matriks koefisien *MFCC* yang mempresentasikan karakteristik spectral sinyal suara. Untuk memperoleh representasi fitur yang lebih ringkas, dilakukan perhitungan rata-rata pada masing-masing koefisien *MFCC* sehingga diperoleh 13 nilai rata-rata *MFCC* untuk setiap sampel suara. Nilai-nilai tersebut digunakan sebagai alat dasar analisis karakteristik suara laki laki dan Perempuan pada penelitian ini.

2.3. Ekstraksi Spectral Centroid

Ekstraksi fitur *Spectral Centroid* dilakukan terhadap seluruh sampel suara menggunakan pustaka Librosa pada bahasa pemrograman Python. Pada tahap awal sinyal suara terlebih dahulu dibagi menjadi beberapa frame menggunakan teknik framing dan diberikan Hamming Window untuk mengurangi efek diskontinuitas pada batas frame. Setelah proses framing dan windowing, nilai *Spectral Centroid* dihitung untuk setiap frame sinyal audio. *Spectral Centroid* merupakan fitur yang digunakan untuk menggambarkan pusat distribusi energi pada spektrum frekuensi suatu sinyal audio. Nilai yang lebih tinggi menunjukkan dominasi energi pada frekuensi tinggi, sedangkan nilai yang lebih rendah menunjukkan dominasi energi pada frekuensi rendah.

Hasil ekstraksi berupa sekumpulan nilai *Spectral Centroid* untuk setiap frame. Selanjutnya dilakukan perhitungan nilai rata-rata (mean) dan standar deviasi dari seluruh frame sebagai representasi karakteristik suara pada setiap sampel audio. Nilai rata-rata digunakan untuk menggambarkan kecenderungan pusat frekuensi suara, sedangkan standar deviasi digunakan untuk menunjukkan tingkat variasi distribusi frekuensi dalam sinyal suara.

2.4. Perhitungan Statistik

Setelah proses ekstraksi ciri dilakukan, tahap selanjutnya adalah perhitungan statistik untuk memperoleh nilai representatif dari setiap fitur yang dihasilkan. Pada fitur *MFCC*, dilakukan perhitungan nilai rata-rata pada setiap koefisien *MFCC* terhadap seluruh frame dalam satu sampel audio. Perhitungan nilai rata-rata digunakan untuk menghasilkan 13 nilai *MFCC* yang merepresentasikan karakteristik suara dari setiap sampel.

Nilai rata-rata (mean) dihitung menggunakan persamaan berikut:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N x_{ij}$$

Keterangan:

μ_i = nilai rata-rata (*mean*) koefisien *MFCC* ke-*i*
 x_{ij} = nilai koefisien *MFCC* ke-*i* pada *frame* ke-*j*
N = jumlah *frame* dalam satu sampel audio
i = indeks koefisien *MFCC* (1 sampai 13)
j = indeks *frame* audio

Pada fitur *Spectral Centroid*, dilakukan perhitungan nilai rata-rata dan standar deviasi dari seluruh nilai *Spectral Centroid* pada setiap *frame* audio. Nilai rata-rata digunakan untuk mengetahui kecenderungan pusat distribusi energi frekuensi, sedangkan standar deviasi digunakan untuk mengetahui tingkat variasi nilai *Spectral Centroid* dalam suatu sampel suara.

Standar deviasi dihitung menggunakan persamaan berikut:

$$s = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (SC_j - \mu)^2}$$

Keterangan:

s = standar deviasi sampel *Spectral Centroid*
SC_j = nilai *Spectral Centroid* pada *frame* ke-*j*
 μ = nilai rata-rata *Spectral Centroid*
N = jumlah *frame* dalam satu sampel audio
N - 1 = faktor koreksi untuk standar deviasi sampel

Nilai statistik yang diperoleh kemudian digunakan untuk membandingkan karakteristik suara laki-laki dan perempuan berdasarkan fitur *MFCC* dan *Spectral Centroid*.

2.5. Analisis Hasil

Tahap analisis dilakukan terhadap hasil ekstraksi ciri *MFCC* dan *Spectral Centroid* yang telah diperoleh dari masing-masing sampel suara. Analisis dilakukan dengan membandingkan nilai statistik yang dihasilkan dari kelompok suara laki-laki dan perempuan. Pada fitur *MFCC*, analisis dilakukan terhadap 13 nilai rata-rata koefisien *MFCC* untuk melihat perbedaan karakteristik spektral yang terdapat pada masing-masing kelompok suara. Sedangkan pada fitur *Spectral Centroid*, analisis dilakukan berdasarkan nilai rata-rata pusat distribusi energi frekuensi serta nilai standar deviasi untuk mengetahui kecenderungan dan variasi distribusi frekuensi pada sinyal suara. Hasil nilai statistik yang diperoleh selanjutnya digunakan untuk mengamati perbedaan karakteristik suara laki-laki dan perempuan berdasarkan fitur *MFCC* dan *Spectral Centroid*.

3. Hasil dan Diskusi

3.1. Hasil Ekstraksi Ciri Menggunakan MFCC

Proses ekstraksi ciri menggunakan metode *Mel-Frequency Cepstral Coefficients (MFCC)* menghasilkan 13 nilai koefisien *MFCC* pada setiap sampel suara yang diucapkan oleh pengguna. Hasil ekstraksi ciri suara dengan sampel salah satu pengguna ditunjukkan pada Tabel 1.

Table 1. Contoh Hasil Ekstraksi Ciri dengan Sampel Salah Satu Pengguna

Sampel	MFCC-1	MFCC-2	MFCC-3	MFCC-4	MFCC-5	MFCC-6	MFCC-7	MFCC-8	MFCC-9	MFCC-10	MFCC-11	MFCC-12	MFCC-13	Rata-rata
1	-398,41	95,97	-4,75	9,47	7,9	16,64	2,41	19,05	2,99	4,61	-2,36	2,99	-1,85	-18,8723
2	-428,46	84,77	-3,33	14,92	10,58	16,99	2,45	15,82	1,19	5,19	0,38	4,83	-0,97	-21,2031
3	-409,6	110,24	-8,6	20,15	16,21	16,77	3,88	21,92	0,49	7,4	-0,54	3,27	-0,52	-16,8408
4	-362,63	99,09	-3,84	30,15	14,84	22,45	1,84	14,42	5,43	7,4	1,31	5,96	0,02	-12,5815
5	-373,99	100,36	-3,54	38,92	11,12	19,55	15,17	12,73	-0,08	10,22	-4,3	8,77	-2,43	-12,8846

Berdasarkan data hasil ekstraksi ciri *MFCC* dengan 13 nilai koefisien tersebut, data sampel dibagi menjadi dua kelompok gender berdasarkan, kelompok suara Perempuan yang diwakili oleh P1 dengan 5 sampel dan P2 dengan 5 sampel, lalu kelompok suara Laki-laki yang diwakili oleh L1 dengan 5 sampel dan L2 dengan 5 sampel.

Untuk mengetahui pola matematis dan karakteristik spectral suara, dilakukan analisis terhadap nilai rata-rata (*mean*) dari 13 koefisien *Mel-Frequency Cepstral Coefficients (MFCC)*. Perhitungan nilai rata-rata dilakukan pada setiap koefisien *MFCC* untuk memperoleh representasi karakteristik spektral dari masing-masing kelompok gender. Selanjutnya, dilakukan pengujian signifikansi statistik menggunakan metode Independent Samples t-Test untuk mengetahui apakah terdapat perbedaan nilai *MFCC* yang signifikan antara kedua kelompok.

Pengujian *Independent Samples t-Test* dilakukan dengan menggunakan dua hipotesis, yaitu H_0 (Hipotesis Nol) yang menyatakan bahwa tidak terdapat perbedaan rata-rata nilai *MFCC* yang signifikan antara kelompok gender, dan H_1 (Hipotesis Alternatif) yang menyatakan bahwa terdapat perbedaan rata-rata nilai *MFCC* yang signifikan antara kedua kelompok gender. Keputusan pengujian ditentukan berdasarkan nilai *p-value* yang diperoleh. Nilai *p-value* menunjukkan tingkat probabilitas untuk menerima hipotesis nol (H_0). Pada penelitian ini, digunakan tingkat signifikansi (*significance level*) sebesar 5% atau $\alpha = 0,05$. Apabila nilai *p-value* $< 0,05$, maka H_0 ditolak sehingga terdapat perbedaan yang signifikan antara kedua kelompok. Sebaliknya, apabila nilai *p-value* $\geq 0,05$, maka H_0 diterima yang menunjukkan bahwa perbedaan nilai *MFCC* antar-kelompok tidak signifikan secara statistik. Hasil perhitungan nilai rata-rata 13 koefisien *MFCC* dan nilai signifikansi (*p-value*) ditunjukkan pada Tabel 2.

Table 2. Hasil Perhitungan Nilai Rata-rata 13 Koefisien MFCC dan Nilai Signifikansi (p-value)

Fitur	Rata-rata Perempuan	Rata-rata Laki laki	Selisih (P-L)	Nilai -p (p-value)
MFCC-1	-544,955	-447,821	-97,134	0,0723
MFCC-2	53,125	98,099	-44,974	0,0154
MFCC-3	-4,495	-7,816	3,321	0,5092
MFCC-4	2,076	-11,679	13,755	0,0926
MFCC-5	11,403	-1,088	12,491	0,0080
MFCC-6	7,578	1,774	5,804	0,1766
MFCC-7	5,109	-2,679	7,788	0,0014
MFCC-8	4,57	-0,884	5,454	0,2775
MFCC-9	1,608	5,248	-3,64	0,2066
MFCC-10	-1,416	3,577	-4,993	0,1384
MFCC-11	-5,423	-10,36	4,937	0,1283
MFCC-12	3,178	-8,667	11,845	0,0000
MFCC-13	-2,738	-12,629	9,891	0,0001

Berdasarkan hasil visualisasi distribusi nilai *MFCC* dan pengujian statistik menggunakan Independent Samples t-Test, ditemukan beberapa koefisien *MFCC* yang menunjukkan perbedaan karakteristik antara sampel suara laki-laki dan perempuan. Perbedaan tersebut menunjukkan adanya variasi pola spektral yang berkaitan dengan karakteristik fisiologis dan resonansi suara pada masing-masing kelompok.

Pada koefisien MFCC-2, diperoleh perbedaan rata-rata yang cukup terlihat antara kelompok laki-laki dan perempuan. Nilai rata-rata MFCC-2 pada suara laki-laki sebesar 98,099, sedangkan pada suara perempuan

sebesar 53,125 dengan nilai signifikansi $p = 0,0154 (< 0,05)$. Hasil ini menunjukkan bahwa terdapat perbedaan yang signifikan pada karakteristik MFCC-2 antara suara laki-laki dan perempuan. Nilai MFCC-2 yang lebih tinggi pada suara laki-laki menunjukkan adanya perbedaan pola distribusi energi spektral pada koefisien tersebut dibandingkan dengan suara perempuan. Perbedaan ini dapat dipengaruhi oleh karakteristik fisiologis pembentukan suara, seperti perbedaan ukuran dan ketebalan pita suara serta panjang saluran vokal yang menyebabkan perbedaan karakteristik akustik antara laki-laki dan perempuan [15]. Dengan demikian, variasi nilai koefisien MFCC dapat menggambarkan adanya perbedaan pola spektral yang menjadi ciri khas masing-masing kelompok suara.

Selain itu, perbedaan karakteristik juga terlihat pada koefisien MFCC-5 dan MFCC-7. Kedua koefisien tersebut menunjukkan pola nilai rata-rata yang berbeda antara kelompok laki-laki dan perempuan, di mana kelompok perempuan memiliki nilai positif (MFCC-5 = 11,403 dan MFCC-7 = 5,109), sedangkan kelompok laki-laki memiliki nilai negatif (MFCC-5 = -1,088 dan MFCC-7 = -2,679). Perbedaan pola nilai ini menunjukkan adanya variasi karakteristik spektral pada kedua kelompok suara. Koefisien MFCC orde menengah berkaitan dengan bentuk *spectral envelope* yang menggambarkan distribusi energi pada spektrum suara, sehingga dapat memberikan informasi mengenai karakteristik resonansi dan warna suara (*timbre*) [16]. Perbedaan struktur saluran vokal antara laki-laki dan perempuan menyebabkan variasi pola resonansi, di mana saluran vokal perempuan yang cenderung lebih pendek menghasilkan pergeseran resonansi ke frekuensi yang lebih tinggi. Hal tersebut menyebabkan adanya perbedaan karakteristik timbre antara suara laki-laki dan perempuan.

Perbedaan paling signifikan ditemukan pada koefisien MFCC-12 dan MFCC-13. Hasil pengujian statistik menunjukkan bahwa kedua koefisien tersebut memiliki nilai signifikansi yang sangat kecil, yaitu kurang dari 0,05, sehingga dapat dinyatakan terdapat perbedaan yang signifikan antara kelompok suara laki-laki dan perempuan. Pada MFCC-12, kelompok perempuan memiliki nilai rata-rata sebesar 3,178, sedangkan kelompok laki-laki sebesar -8,667. Sementara itu, pada MFCC-13 nilai rata-rata kelompok perempuan sebesar -2,738 dan kelompok laki-laki sebesar -12,629. Perbedaan nilai yang cukup besar pada kedua koefisien ini menunjukkan adanya variasi karakteristik spektral antara kedua kelompok suara. Koefisien MFCC orde tinggi berperan dalam menangkap detail perubahan spektrum yang lebih halus, termasuk variasi komponen harmonik suara (Terasawa dkk., 2012).

Meskipun MFCC-1 menunjukkan perbedaan nilai rata-rata yang cukup besar antara perempuan (-544,955) dan laki-laki (-447,821), hasil pengujian statistik menunjukkan bahwa perbedaan tersebut tidak signifikan dengan nilai $p = 0,0723 (> 0,05)$. Hal ini menunjukkan bahwa MFCC-1 belum dapat dijadikan parameter pembeda gender secara kuat. MFCC-1 lebih berkaitan dengan informasi energi sinyal atau amplitudo rekaman, sehingga nilainya dapat dipengaruhi oleh faktor eksternal seperti jarak mikrofon, tingkat volume suara, dan noise lingkungan. Variasi kondisi perekaman menyebabkan penyebaran nilai yang lebih besar dalam kelompok yang sama sehingga perbedaannya tidak menunjukkan signifikansi statistik.

3.2. Hasil Ekstraksi Ciri Menggunakan Spectral Centroid

Hasil ekstraksi fitur *Spectral Centroid* menghasilkan nilai rata-rata pusat distribusi energi frekuensi dan nilai standar deviasi pada setiap sampel suara. Nilai rata-rata *Spectral Centroid* menunjukkan posisi pusat frekuensi dari energi sinyal audio, sedangkan standar deviasi menunjukkan tingkat perubahan distribusi frekuensi selama proses pengucapan berlangsung.

Table 3. Nilai Rata-rata Spectral Centroid dan Standar Deviasi

Responden	Gender	Rerata Spectral Centroid (Hz)	Rerata Standar Deviasi (Hz)
P1	Perempuan	1363,36	389,72
P2	Perempuan	4847,39	1837,21
Rata-rata Kelompok		3105,38	1113,47
L1	Laki Laki	2400,07	714,77
L2	Laki Laki	1489,74	258,82
Rata-rata Kelompok		1944,91	486,80

Berdasarkan Tabel 3, kelompok suara perempuan memiliki nilai rata-rata *Spectral Centroid* sebesar 3105,38 Hz, sedangkan kelompok suara laki-laki memiliki nilai rata-rata sebesar 1944,91 Hz. Hasil tersebut menunjukkan

bahwa suara perempuan memiliki pusat distribusi energi frekuensi yang lebih tinggi dibandingkan suara laki-laki.

Perbedaan nilai *Spectral Centroid* ini dapat dikaitkan dengan karakteristik fisiologis pembentukan suara. Secara umum, perempuan memiliki ukuran pita suara yang lebih pendek dan tipis serta panjang saluran vokal yang lebih kecil dibandingkan laki-laki. Perbedaan anatomi tersebut menyebabkan frekuensi fundamental (F_0) dan beberapa komponen resonansi suara perempuan cenderung berada pada rentang frekuensi yang lebih tinggi. Karena *Spectral Centroid* menggambarkan posisi rata-rata energi dalam spektrum, pergeseran energi menuju frekuensi tinggi menyebabkan nilai *Spectral Centroid* menjadi lebih besar. Hal ini sesuai dengan karakteristik persepsi suara perempuan yang umumnya terdengar lebih terang (*bright*) dibandingkan suara laki-laki.

Selain nilai rata-rata, variasi nilai *Spectral Centroid* juga dapat diamati melalui standar deviasi. Kelompok perempuan memiliki nilai rata-rata standar deviasi sebesar 1113,47 Hz, sedangkan kelompok laki-laki sebesar 486,80 Hz. Nilai standar deviasi yang lebih tinggi menunjukkan adanya perubahan distribusi energi frekuensi yang lebih besar selama proses pengucapan.

Perbedaan variabilitas tersebut menunjukkan bahwa karakteristik spektral suara perempuan pada data penelitian ini mengalami fluktuasi yang lebih besar dibandingkan suara laki-laki. Sebaliknya, suara laki-laki menunjukkan pola distribusi energi yang relatif lebih stabil dengan variasi frekuensi yang lebih rendah. Hal ini menunjukkan bahwa standar deviasi *Spectral Centroid* dapat memberikan informasi tambahan mengenai dinamika perubahan karakteristik spektrum suara pada masing-masing kelompok.

3.3. Perbandingan Karakteristik Suara Berdasarkan Gender

Berdasarkan hasil ekstraksi ciri menggunakan *MFCC* dan *Spectral Centroid*, diperoleh adanya perbedaan karakteristik fitur antara kelompok suara laki-laki dan perempuan. Perbedaan tersebut terlihat dari variasi nilai koefisien *MFCC* serta nilai rata-rata *Spectral Centroid* yang dihasilkan pada masing-masing sampel suara.

Pada fitur *MFCC*, perbedaan nilai antar kelompok menunjukkan adanya variasi pola distribusi spektral pada sinyal audio. *MFCC* mampu merepresentasikan karakteristik suara melalui koefisien yang diperoleh dari proses transformasi spektrum menggunakan skala Mel. Perbedaan nilai koefisien tersebut menunjukkan adanya perbedaan karakteristik akustik pada data suara yang dianalisis, meskipun terdapat beberapa koefisien dengan nilai yang relatif berdekatan antar sampel.

Sementara itu, hasil ekstraksi *Spectral Centroid* menunjukkan adanya kecenderungan perbedaan pusat distribusi energi frekuensi antara suara laki-laki dan perempuan. Nilai *Spectral Centroid* yang lebih tinggi menunjukkan bahwa energi sinyal lebih banyak terkonsentrasi pada komponen frekuensi yang lebih tinggi. Perbedaan tersebut dapat berkaitan dengan karakteristik fisiologis produksi suara, seperti perbedaan ukuran dan ketebalan pita suara serta panjang saluran vokal yang memengaruhi karakteristik frekuensi yang dihasilkan.

Namun, hasil penelitian juga menunjukkan adanya variasi nilai antar individu dalam kelompok gender yang sama. Variasi tersebut dapat dipengaruhi oleh faktor individu seperti karakteristik vokal, cara pengucapan, serta kondisi lingkungan saat proses perekaman. Oleh karena itu, fitur *MFCC* dan *Spectral Centroid* pada penelitian ini digunakan sebagai representasi karakteristik suara, bukan sebagai metode klasifikasi gender.

4. Kesimpulan

Berdasarkan hasil penelitian mengenai ekstraksi ciri suara menggunakan metode Mel-Frequency Cepstral Coefficients (*MFCC*) dan *Spectral Centroid*, dapat disimpulkan bahwa kedua fitur tersebut dapat digunakan untuk merepresentasikan karakteristik spektral pada sinyal suara laki-laki dan perempuan. Hasil analisis menggunakan *MFCC* menunjukkan adanya perbedaan nilai pada beberapa koefisien antara kedua kelompok suara. Berdasarkan pengujian statistik menggunakan *Independent Samples t-Test*, beberapa koefisien *MFCC* memiliki perbedaan yang signifikan dengan nilai $p\text{-value} < 0,05$, yaitu pada *MFCC*-2, *MFCC*-5, *MFCC*-7, *MFCC*-12, dan *MFCC*-13. Hasil tersebut menunjukkan adanya variasi pola spektral pada sinyal suara dari kedua kelompok. Pada fitur *Spectral Centroid*, kelompok suara perempuan menunjukkan nilai rata-rata yang lebih tinggi dibandingkan kelompok suara laki-laki. Hal ini menunjukkan adanya kecenderungan distribusi energi frekuensi yang lebih tinggi pada suara perempuan. Selain itu, nilai standar deviasi *Spectral Centroid*

menunjukkan adanya perbedaan tingkat variasi distribusi frekuensi antar kelompok suara. Meskipun terdapat perbedaan karakteristik fitur antara kelompok suara laki-laki dan perempuan, penelitian ini masih terbatas pada analisis representasi karakteristik suara dengan jumlah sampel yang terbatas. Oleh karena itu, fitur *MFCC* dan *Spectral Centroid* pada penelitian ini digunakan untuk menggambarkan perbedaan pola akustik suara dan belum digunakan sebagai metode klasifikasi gender. Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan untuk menambah jumlah responden serta memperluas variasi fonem atau kata yang diucapkan agar dataset menjadi lebih beragam. Selain itu, rentang usia responden juga perlu diperluas untuk mengamati pengaruh perkembangan biologis pita suara terhadap karakteristik spektral yang dihasilkan. Penelitian selanjutnya juga disarankan untuk melakukan proses perekaman pada lingkungan yang terkondisi, seperti ruangan kedap suara (*soundproof*), dengan jarak mikrofon yang tetap dan konsisten. Langkah tersebut diharapkan dapat meminimalkan interferensi *noise* lingkungan, menjaga konsistensi amplitudo sinyal audio, serta meningkatkan keandalan hasil ekstraksi fitur, khususnya pada koefisien *MFCC* yang tidak menunjukkan perbedaan signifikan. Berdasarkan hasil penelitian mengenai ekstraksi ciri suara menggunakan metode Mel-Frequency Cepstral Coefficients (*MFCC*) dan *Spectral Centroid*, dapat disimpulkan bahwa kedua fitur tersebut dapat digunakan untuk merepresentasikan karakteristik spektral pada sinyal suara laki-laki dan perempuan. Hasil analisis menggunakan *MFCC* menunjukkan adanya perbedaan nilai pada beberapa koefisien antara kedua kelompok suara. Berdasarkan pengujian statistik menggunakan Independent Samples t-Test, beberapa koefisien *MFCC* memiliki perbedaan yang signifikan dengan nilai p-value < 0,05, yaitu pada MFCC-2, MFCC-5, MFCC-7, MFCC-12, dan MFCC-13. Hasil tersebut menunjukkan adanya variasi pola spektral pada sinyal suara dari kedua kelompok. Pada fitur *Spectral Centroid*, kelompok suara perempuan menunjukkan nilai rata-rata yang lebih tinggi dibandingkan kelompok suara laki-laki. Hal ini menunjukkan adanya kecenderungan distribusi energi frekuensi yang lebih tinggi pada suara perempuan. Selain itu, nilai standar deviasi *Spectral Centroid* menunjukkan adanya perbedaan tingkat variasi distribusi frekuensi antar kelompok suara. Meskipun terdapat perbedaan karakteristik fitur antara kelompok suara laki-laki dan perempuan, penelitian ini masih terbatas pada analisis representasi karakteristik suara dengan jumlah sampel yang terbatas. Oleh karena itu, fitur *MFCC* dan *Spectral Centroid* pada penelitian ini digunakan untuk menggambarkan perbedaan pola akustik suara dan belum digunakan sebagai metode klasifikasi gender.

Referensi

- [1] H. Hermansky, 'Coding and decoding of messages in human speech communication: Implications for machine recognition of speech', *Speech Commun.*, vol. 106, pp. 112–117, Jan. 2019, doi: 10.1016/J.SPECOM.2018.12.004.
- [2] M. Labied and A. Belangour, 'Automatic Speech Recognition Features Extraction Techniques: A Multi-criteria Comparison', *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 8, pp. 177–182, 2021, doi: 10.14569/IJACSA.2021.0120821.
- [3] T. Giannakopoulos and A. Pikrakis, *Introduction to Audio Analysis: A MATLAB Approach*. 2014. doi: 10.1016/C2012-0-03524-7.
- [4] A. A. Shafah, P. P. Adikara, and S. Adinugroho, 'Klasifikasi Jenis Kelamin Berdasarkan Suara Menggunakan Metode Learning Vector Quantization', *J. Pengemb. Teknol. dan Ilmu Komput.*, vol. 4, no. 7, pp. 2301–2308, Aug. 2020.
- [5] P. Negeri and L. Rahmadani, 'KLASIFIKASI GENDER BERDASARKAN SUARA DENGAN NAIVE BAYES DAN MEL FREQUENCY CEPSTRAL COEFFICIENT', *VOCATECH Vocat. Educ. Technol. J.*, vol. 2, no. 1, pp. 19–26, Oct. 2020, doi: 10.38038/VOCATECH.V2I1.45.
- [6] D. T. Kusuma, 'Fast Fourier Transform (FFT) Dalam Transformasi Sinyal Frekuensi Suara Sebagai Upaya Perolehan Average Energy (AE) Musik', *Petir*, vol. 14, no. 1, pp. 28–35, 2020, doi: 10.33322/petir.v14i1.1022.
- [7] R. A. N. Diaz, N. L. G. P. Suwirmayanti, and K. Budiarta, 'Perbandingan Kualitas Pengenalan Suara Untuk Ekstraksi Fitur Menggunakan Mfcc Dan Spectral', *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 6, no. 1, pp. 58–63, 2024, doi: 10.53580/naratif.v6i1.281.
- [8] 'Spectral centroid - Wikipedia'.
- [9] R. Umar, I. Riadi, and A. Hanif, 'Analisis Bentuk Pola Suara Menggunakan Ekstraksi Ciri Mel-Frequency Cepstral Coefficients (MFCC)', *Cogito To Smart J.*, vol. 4, no. 2, pp. 294–304, 2019, doi: 10.31154/cogito.v4i2.130.294-304.
- [10] H. (Heriyanto) Heriyanto, S. (Sri) Hartati, and A. E. (Agfianto) Putra, 'Ekstraksi Ciri Mel Frequency Cepstral Coefficient (Mfcc) Dan Rerata Coefficient Untuk Pengecekan Bacaan Al-qur'an', *Telematika*, vol. 15, no. 2, pp. 99–108, Oct. 2018, doi: 10.31315/TELEMATIKA.V15I2.3123.
- [11] S. Kazazis, P. Depalle, and S. McAdams, 'Interval and Ratio Scaling of Spectral Audio Descriptors', *Front. Psychol.*, vol. 13, p. 835401, Mar. 2022, doi: 10.3389/FPSYG.2022.835401/TEXT.
- [12] J. Delisle, 'Spectral centroid — Timbre and Orchestration Resource'.
- [13] V. V. Bilqis, A. W. Septiani, F. N. Zildjian, and F. D. Ratnasari, 'Analisis Perbedaan Spektrum Frekuensi Sinyal Suara Vokal Manusia Menggunakan Fast Fourier Transform (FFT) Berbasis Python', *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 5, no. 2, pp. 3548–3558, May 2026, doi: 10.31004/RIGGS.V5I2.8994.
- [14] A. A. Kasim, M. Bakri, I. Mahmudi, R. Rahmawati, and Z. Zulnabil, 'Artificial Intelligent for Human Emotion Detection with the Mel-Frequency Cepstral Coefficient (MFCC)', *JUITA J. Inform.*, vol. 11, no. 1, pp. 47–55, May 2023, doi: 10.30595/JUITA.V11I1.15435.
- [15] 'What Are Your Vocal Cords? (n.d.)'. Retrieved June 27, 2026, from <https://my.clevelandclinic.org/health/body/24456-vocal-cords>
- [16] H. Terasawa, J. Berger, and S. Makino, 'In search of a perceptual metric for timbre: Dissimilarity judgments among synthetic sounds with MFCC-derived spectral envelopes', *AES J. Audio Eng. Soc.*, vol. 60, no. 9, pp. 674–685, 2012.