

Identifikasi Nada Dasar Kendang Menggunakan MFCC dan KNN

Ni Kadek Yulia Dewi^{a1}, I Ketut Gede Suhartana^{a2}

^{a1}Informatics Department, Udayana University
Bali, Indonesia
¹yuldedew1104@gmail.com
²ikg.suhartana@unud.ac.id

Abstract

Balinese kendang is a Balinese musical instrument that is closely related to the art of karawitan. Usually drums are played in a musical instrument show in Bali. Balinese drums are played in pairs, which consist of lanang drums and wadon drums. The sound features used in this system are extracted from the MFCC algorithm which are then classified using the KNN algorithm. The results of the system show the best classification results with an accuracy of 90% with parameter $K = 1$ and can correctly recognize 54 tones out of 60 tones.

Keywords: Basic Tone, Kendang, Music, Classification, Audio Extraction

1. Pendahuluan

Kendang Bali adalah instrumen musik bali yang sangat berhubungan dengan kesenian karawitan. Biasanya alat musik kendang dimainkan dalam sebuah pertunjukan alat musik di bali. Kendang bali dimainkan secara berpasangan, yakni terdiri dari kendang lanang dan kendang wadon. Dalam mempelajari alat musik kendang dibutuhkan pelatih yang berpengalaman untuk mengecek apakah suara yang dihasilkan sudah benar atau tidak. Namun kurangnya tenaga pelatih kendang membuat sebagian orang enggan untuk mempelajari alat musik ini. Dengan permasalahan tersebut, peran aplikasi pembelajaran sangatlah penting untuk menunjang pembelajaran serta tetap melestarikan kesenian ini sehingga seseorang tetap bisa mempelajari alat musik tersebut tanpa didampingi seorang pelatih.

Beberapa penelitian serupa dengan masalah ini adalah "Voice Recognition untuk Sistem Keamanan PC Menggunakan Metode MFCC dan DTW". Pada penelitian yang dilakukan oleh P.T handoko et al, yang membahas tentang penggunaan metode MFCC dalam keamanan sistem PC untuk mengamankan kinerja, data, fungsi atau proses komputer dengan teknologi biometrik atau teknologi berbasis ukuran pada manusia seperti sidik jari, wajah, kornea mata, dan lain-lain dapat mengenali suara yang memiliki otoritas dengan kriteria dalam keadaan noise 82% dan hening 86% [1]. Selain itu penelitian "Deteksi Nada Tunggal Alat Musik Kecapi Bugis Makassar Menggunakan Metode Mel Frequency Cepstral Coefficient (MFCC) Dan Klasifikasi K-Nearest Neighbour". Penelitian ini dibuat untuk alat musik kecap sering digunakan dalam festival musik Sulawesi di berbagai daerah, namun sering terdapat permasalahan pada saat penyetalan alat musik kecap karena membutuhkan waktu cukup lama sehingga dibuat sistem yang dapat mengidentifikasi nada yang terdapat pada alat musik kecap melalui pengolahan suara. Hasil dari penelitian ini adalah akurasi sistem yang paling terbaik didapatkan ketika nilai windowing 120 dengan akurasi 81.42%. Hal ini disebabkan karena windowing bertujuan untuk mengurangi efek diskontinuitas pada ujung-ujung frame yang dihasilkan oleh proses framing. Sistem yang terbaik didapatkan ketika menggunakan overlap hal ini disebabkan karena pada saat non-overlap terjadi aliasing sehingga nada luaran lebih cepat dari pada nada aslinya. Dan pada akurasi tertinggi jenis KNN yang digunakan yaitu distance dengan variabel $k=1$ [2].

Menguji menggunakan data training berupa nada dasar Kendang yang memiliki nada dasar yang benar yang akan melalui proses ekstraksi fitur menggunakan metode Mel-Frequency Cepstral Coefficients (MFCC). Hasil dari ekstraksi fitur ini akan disimpan ke dalam database. Sedangkan untuk data uji akan melalui proses ekstraksi fitur yang sama dengan metode MFCC, kemudian fitur hasil ekstraksi akan melalui proses klasifikasi menggunakan metode K-Nearest Neighbor (KNN) untuk mengetahui apakah nada dasar kendang sudah benar atau belum.

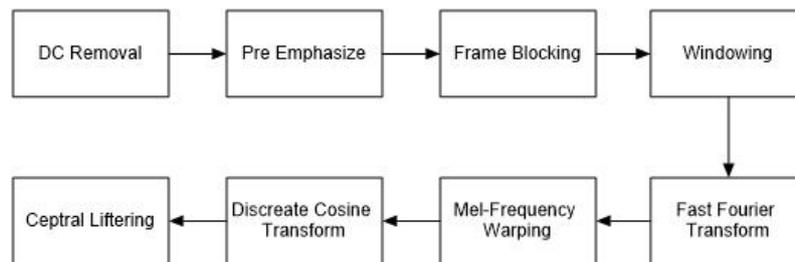
Adapun hasil yang diharapkan adalah tingginya tingkat akurasi pada hasil pengelompokan sehingga pendekatan metode ekstraksi fitur yang digunakan pada penelitian ini dapat digunakan sebagai metode alternatif untuk ekstraksi fitur dan klasifikasi nada dasar alat musik Kendang.

2. Metode Penelitian

Metode yang digunakan pada penelitian ini yaitu *Mel Frequency Cepstral Coefficients* (MFCC) untuk ekstraksi ciri dan *K-Nearest Neighbor* untuk klasifikasi nada.

2.1. Mel Frequency Cepstral Coefficients (MFCC)

Mel Frequency Cepstrum Coefficients adalah metode yang digunakan untuk melakukan ekstraksi fitur untuk mendapatkan suatu parameter dan informasi mengenai ciri dari suatu sinyal. MFCC merupakan cara yang paling sering digunakan untuk ekstraksi fitur audio signal, karena menggunakan komputasi logaritmik sesuai dengan lingkup pendengaran manusia. Dimana sinyal suara akan di filter secara linear untuk frekuensi rendah (dibawah 1000 Hz) dan secara logaritmik untuk frekuensi tinggi (diatas 1000 Hz). Sehingga dapat merepresentasikan parameter suara dengan baik [4]. Beberapa keunggulan menggunakan MFCC sebagai ekstraksi fitur suara adalah mampu untuk menangkap karakteristik suara dan informasi-informasi penting dalam sinyal suara, menghasilkan data seminimal mungkin tanpa menghilangkan informasi-informasi penting yang ada, dan mengadaptasi organ pendengaran manusia dalam melakukan persepsi terhadap sinyal suara [3]. Diagram MFCC dapat dilihat pada gambar 1.



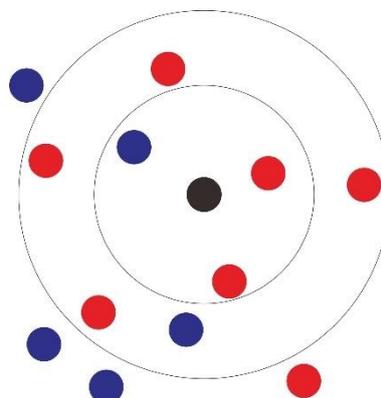
Gambar 1. Diagram MFCC

Tahap pertama dalam ekstraksi fitur menggunakan MFCC adalah *DC Removal* yang bertujuan untuk membuang data-data yang tidak dibutuhkan di dalam proses ekstraksi. Selanjutnya adalah proses *Pre-emphasize* bertujuan untuk mempertahankan frekuensi frekuensi tinggi pada sebuah spektrum yang umumnya tereliminasi pada saat proses produksi suara, selain itu *pre-emphasize* dapat mengurangi noise pada suara masukan sehingga tingkat akurasi dari proses ekstraksi ciri dapat ditingkatkan. Langkah selanjutnya adalah proses *Frame Blocking* membagi sinyal suara ke dalam *frame-frame* dengan waktu tertentu yang lebih singkat. Setelah itu *frame-frame* tersebut akan dibawa ke proses *Windowing* dimana proses ini berfungsi untuk meminimalisir diskontinuitas pada permulaan dan akhir *frame*. Proses selanjutnya adalah *Fast Fourier Transform* untuk mengonversi setiap *frame* N sampel dari domain waktu ke domain frekuensi. Kemudian masuk ke tahap *Mel-Frequency Warping* untuk melakukan filter pada sinyal untuk setiap *frame* menggunakan *Mel Filterbank* yang terdiri dari *triangular window* sebanyak N. Selanjutnya adalah tahap *Discrete Cosine Transform* untuk mengonversi nilai mel kembali kedalam domain waktu. Dan proses terakhir adalah *Cepstral Lifting* yang berfungsi untuk menghaluskan spektrum hasil dari *main processor* sehingga dapat digunakan lebih baik untuk *pattern matching*.

2.2. K Nearest Neighbor (KNN)

K-Nearest Neighbor merupakan salah satu algoritma pengklasifikasian yang cukup mudah dipahami, karena KNN mencari jumlah kesamaan terbanyak antara data yang diuji dengan data latih. Data uji tersebut akan masuk ke dalam kelas dengan jumlah kesamaan terbanyak. Konsep dasar dari *K-Nearest Neighbor* adalah seperti pada algoritma *Nearest Neighbor*, yaitu mencari jarak terdekat dari nilai yang akan dievaluasi (titik queri) dengan tetangga terdekatnya dalam suatu data [2]. Kelebihan KNN menggunakan prinsip yang sederhana, bekerja berdasarkan jarak terdekat dari sampel uji ke

sampel latih dan tidak memperhitungkan kemungkinan distribusi dari masing-masing kelas [5]. Ilustrasi KNN dapat dilihat pada gambar 2.



Gambar 2. Ilustrasi KNN

Langkah pertama algoritma KNN adalah menentukan nilai K sebagai jumlah mayoritas untuk proses klasifikasi. Penentuan nilai K sangat penting karena akan sangat mempengaruhi tingkat akurasi dari proses klasifikasi. Setelah menentukan nilai K adalah menghitung jarak data uji ke data latih. Ada beberapa rumus atau persamaan untuk menghitung jarak seperti *Manhattan Distance*, *Minkowsky Distance*, *Chebyshev Distance*, dan *Euclidean Distance*.

Persamaan untuk menghitung jarak pada penelitian ini menggunakan *Euclidean Distance*. Namun terdapat kelemahan dari fungsi jarak *Euclidean* ini yaitu jika salah satu input atribut memiliki rentang yang relatif besar dapat mengalahkan atribut lainnya. Akibatnya, jarak sering dinormalisasi dengan membagi jarak untuk setiap atribut dengan rentang (yaitu nilai maksimum-nilai minimum) dari atribut sehingga nilai untuk setiap atribut memiliki rentang baru yang dinormalisasi dari 0 hingga 1 [6]. Penerapan dari rumus *Euclidean Distance* adalah dengan mengakarkan nilai dari variabel data latih dikurangi dengan nilai variabel data uji yang sudah di pangkatkan dengan dua. Jika terdapat lebih dari satu variabel, maka akumulasikan pemangkatan dua yang sebelumnya sudah dilakukan pengurangan data latih dikurangi dengan data uji. Setelah itu akan menghasilkan jarak dari data latih dan data uji yang selanjutnya akan ditentukan apakah data testing sudah benar atau tidak dengan menggunakan algoritma KNN. Rumus *Euclidean Distance* dapat dilihat pada persamaan 1.

$$D(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

Keterangan :

$D(x,y)$ = jarak antara data uji dengan data latih

x = data uji

y = data latih

k = variabel data

n = jumlah data latih

3. Hasil dan Diskusi

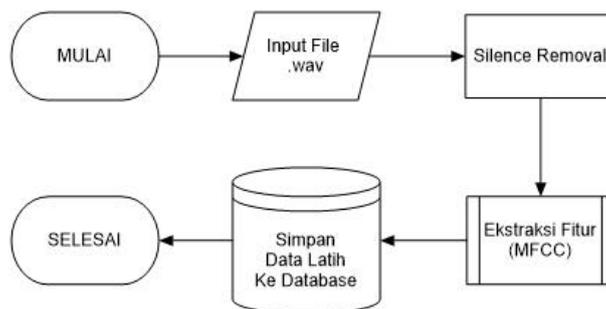
Sistem identifikasi nada ini dimulai dari proses ekstraksi fitur suara menggunakan metode MFCC yang kemudian hasil dari ekstraksi fitur tersebut akan disimpan ke dalam database dan digunakan sebagai data latih, sedangkan untuk data uji sendiri akan dilakukan ekstraksi fitur dengan metode MFCC juga yang kemudian akan dilakukan pengklasifikasian menggunakan metode KNN. Dataset yang digunakan dalam penelitian ini adalah sebanyak 210 data dengan 150 data sebagai data latih, dan 60 data sebagai data uji dengan pembagian nada seperti pada tabel 1.

| Jenis Nada | Data Latih | Data Uji |
|---------------|------------|----------|
| Dum | 50 | 20 |
| Tak / Ka | 50 | 20 |
| Slap | 50 | 20 |
| Jumlah | 150 | 60 |

Tabel 1. Dataset penelitian

3.1. Tahap Pelatihan

Tahap pelatihan adalah tahapan untuk mengolah data latih kemudian menyimpannya ke dalam database untuk dijadikan sebagai referensi untuk tahap *pengujian*. Pada tahap pelatihan, jenis nada akan dikenali sesuai dengan format nama file yang sudah dibuat. Proses ekstraksi fitur menggunakan metode MFCC menggunakan *library python* yaitu *librosa* dengan penentuan parameter panjang *frame* 10 ms dan lebar *window* sebesar 20 ms sehingga overlapping yang terjadi sebesar 50%. Sebelum melalui proses ekstraksi, file suara akan melalui proses *silence removal* dengan pemotongan keheningan pada awal dan akhir sinyal suara dengan nilai desibel dibawah 30 desibel menggunakan fungsi *librosa.effects.trim()*. Penentuan nilai 30 desibel karena suara dengan desibel 30 adalah suara dengan kebisingan rendah seperti suara bisikan dan tergolong sangat sunyi [7], sehingga *silence removal* akan membuang suara dibawah 30 desibel di awal dan akhir rekaman sehingga ekstraksi dilakukan hanya pada suara pukulan alat musik kendang saja. *Flowchart* tahap pelatihan dapat dilihat pada gambar 3.



Gambar 3. Flowchart tahap pelatihan

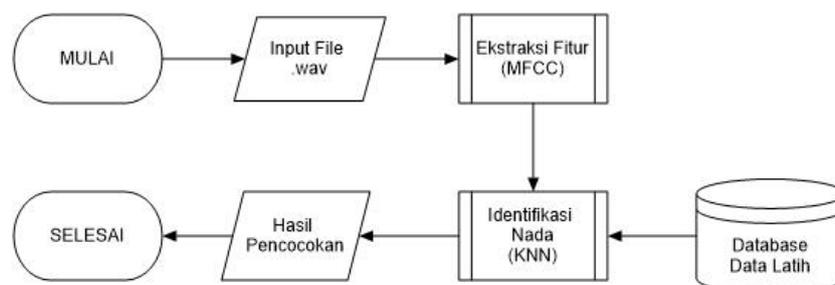
Hasil dari proses ekstraksi didapatkan suatu vektor ciri akan yang berupa *array* dua dimensi yang kemudian akan dicari rata-rata dari tiap koefisien setelah itu akan didapatkan *array* 1 dimensi dengan panjang sesuai jumlah koefisien yang ditentukan. Kemudian *array* tersebut dikonversi ke string karena database yang akan digunakan adalah MySQL yang tidak dapat menyimpan data *array*. Struktur *database* yang digunakan untuk menyimpan data hasil ekstraksi ciri memiliki tiga kolom yaitu kolom "id" sebagai *primary key*, kolom "jenis_nada" untuk menyimpan jenis nada hasil ekstraksi ciri, dan kolom "data" untuk menyimpan hasil ekstraksi ciri data uji. Hasil ekstraksi ciri pada database dapat dilihat pada gambar 4.

| id | jenis_nada | data |
|----|------------|--|
| 1 | dum | [-324.82358 148.69264 36.52158 40.62... |
| 2 | dum | [-269.31342 165.75792 6.789374 29.17806... |
| 3 | dum | [-265.55814 176.32838 2.5927823 12.80517... |
| 4 | dum | [-309.3005 150.56662 21.90566 19.37736... |
| 5 | dum | [-311.9887 158.94077 30.43635 34.20... |

Gambar 4. Hasil ekstraksi di database

3.2. Tahap Pengujian

Tahap pengujian adalah proses untuk menguji apakah data uji sudah benar atau tidak. Proses pengujian akan melalui proses *silence removal* sama seperti proses pelatihan dan ekstraksi fitur menggunakan algoritma MFCC, kemudian hasil ekstraksi akan melalui proses klasifikasi menggunakan algoritma KNN dengan cara mencocokkan antara hasil ekstraksi data uji dengan hasil ekstraksi data latih yang tersimpan di dalam database. Flowchart tahap pengujian dapat dilihat pada gambar 5.



Gambar 5. Flowchart tahap pengujian

Hasil klasifikasi yang dilakukan pada penelitian ini menggunakan 60 data uji dengan 20 data untuk nada dum, 20 data untuk nada tak, dan 20 data untuk nada slap dengan lima parameter K yaitu K=1, K=3, K=5, K=7, K=9 dengan akurasi terbaik pada nilai K=1 dan akurasi terburuk pada nilai K=3. Hasil klasifikasi dapat dilihat pada table 2.

| K | Nada Dum | | Nada Tak | | Nada Slap | | Akurasi |
|---|----------|-------|----------|-------|-----------|-------|---------|
| | Benar | Salah | Benar | Salah | Benar | Salah | |
| 1 | 19 | 1 | 15 | 5 | 20 | 20 | 90% |
| 3 | 11 | 9 | 14 | 6 | 20 | 20 | 75% |
| 5 | 12 | 8 | 14 | 6 | 20 | 20 | 76,6% |
| 7 | 13 | 7 | 14 | 6 | 20 | 20 | 78,33% |
| 9 | 14 | 6 | 14 | 6 | 20 | 20 | 80% |

Tabel 2. Hasil klasifikasi

Hasil klasifikasi terbaik dengan parameter K=1 dengan akurasi sebesar 90% dan berhasil mengklasifikasikan nada dengan benar sebanyak 54 dari 60 nada sedangkan dan akurasi terburuk adalah K = 3 dengan akurasi sebesar 75% yang berhasil mengklasifikasikan nada dengan benar sebanyak 46 dari 60 nada.

4. Kesimpulan

Berdasarkan hasil penelitian dan pengujian sistem yang dilakukan terhadap 20 nada dum, 20 nada tak, dan 20 nada slap pada alat musik kendang dapat disimpulkan bahwa penggunaan algoritma MFCC dan KNN dapat mengklasifikasikan nada alat musik Kendang sampai dengan akurasi 90% dengan K=1. Dengan demikian dapat dikatakan bahwa penggunaan algoritma *Mel-Frequency Cepstral Coefficients* dan *K-Nearest Neighbor* terbukti cukup baik untuk mengklasifikasikan nada dasar alat musik Kendang dan dapat dijadikan sebagai acuan untuk penelitian kedepannya. Kegagalan penelitian sebesar 10% disebabkan karena kondisi lingkungan yang terdapat noise. Beberapa saran untuk mendapatkan akurasi yang lebih baik adalah dengan mengkondisikan tempat yang sedikit terdapat noise dan menggunakan alat perekaman yang baik sehingga kualitas suara yang didapatkan lebih baik.

Referensi

- [1] D. T. Handoko and P. Kasih, "Voice Recognition untuk Sistem Keamanan PC Menggunakan Metode MFCC dan DTW," *Gener. J.*, vol. 2, 2018.
- [2] A. F. Ryamizard, B. Hidayat, and S. Saidah, "Deteksi Nada Tunggal Alat Musik Kecapi Bugis Makassar Menggunakan Metode Mel Frequency Cepstral Coefficient (Mfcc) Dan Klasifikasi K-Nearest Neighbour (Knn)," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 4715–4721, 2018.
- [3] M. Frequency and S. Coefficients, "Metoda Mel Frequency Cepstrum Coefficients (MFCC) untuk Mengenali Ucapan pada Bahasa Indonesia Torkis Nasution," *J. Sains dan Teknol. Inf.*, vol. 1, 2012.
- [4] F. N. Suciani, E. C. Djamal, and R. Ilyas, "Identifikasi Nama Surat Juz Amma dengan Perintah Suara Menggunakan MFCC dan Backpropagation," *Semin. Nas. Apl. Teknol. Inf. 2018*, pp. 18–23, 2018.
- [5] A. Budianto, D. Maryono, and R. Ariyuana, "Perbandingan K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dalam Pengenalan Karakter Plat Kendaraan Bermotor," *J. Ilm. Pendidik. Tek. Kejuru.*, vol. 11, no. 1, 2018.
- [6] A. R. Lubis and M. Lubis, "Optimization of distance formula in K-Nearest Neighbor method," vol. 9, no. 1, pp. 326–338, 2020, doi: 10.11591/eei.v9i1.1464.
- [7] Z. C., "How Loud is a Decibel?," 2018. <https://www.sylvane.com/blog/how-loud-is-a-decibel/> (accessed Sep. 30, 2021).