

Identifikasi Nada Dari Sinyal Suara Alat Musik Instrumen Menggunakan Metode Mel Frequency Cepstrum Coefficients dan Hidden Markov Model

Riyan Firmansyah¹, Esmeralda C Djamal, Rezki Yuniarti

Jurusan Informatika, Fakultas Sains dan Informatika
Universitas Jenderal Achmad Yani
Kota Cimahi

¹riyanfirmansyah048@gmail.com

Abstract — Semakin berkembangnya seni musik mengakibatkan banyaknya peminatan dalam seni musik. Hal mendasar yang harus dipelajari dalam seni musik adalah mengenali nada dasar yang dihasilkan oleh alat musik tersebut salah satunya adalah nada mayor. Namun untuk dapat mengenali nada mayor, seseorang membutuhkan latihan bermusik yang cukup lama serta biaya yang tidak sedikit. Hal tersebut dapat membatasi seseorang dalam mempelajari seni musik. Oleh karena itu, diperlukan identifikasi nada dari alat musik, namun dalam identifikasi tidaklah mudah. Penelitian ini telah membuat sistem identifikasi nada mayor dari perekaman lima alat musik menggunakan 10 kali perulangan dengan durasi perekaman dua detik untuk setiap nada. Sistem dibangun dengan menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) sebagai ekstraksi nada dan diidentifikasi menggunakan Hidden Markov Model (HMM). MFCC digunakan karena memiliki kemiripan dengan sistem pendengaran manusia yang memperhitungkan sifat non linier. Hasil dari MFCC dilakukan pelabelan frame menggunakan K-Means, yang kemudian digunakan sebagai masukan parameter HMM untuk mengidentifikasi 12 nada mayor yang kemudian direpresentasikan dalam bentuk simbol. Hasil pengujian menunjukkan akurasi terhadap data baru sebesar 74% untuk Biola, 81% untuk Ukulele, 61% untuk Gitar, 82% untuk Piano dan 30% untuk Bass. Sedangkan untuk data latih mendapatkan akurasi sebesar 92%.

Kata kunci— *Mel-Frequency Cepstrum Coefficients (MFCC); K-Means; Hidden Markov Model (HMM); Nada.*

I. PENDAHULUAN

Semakin berkembangnya seni musik mengakibatkan banyaknya minat dalam mempelajari musik. Dalam mempelajari musik, hal mendasar yang harus dipelajari adalah mengenali nada yang dihasilkan oleh alat musik tersebut. Namun untuk dapat mengenali nada yang dihasilkan oleh alat musik, seseorang membutuhkan latihan bermusik yang cukup lama serta biaya yang tidak sedikit. Hal tersebut membatasi peminat seni musik dalam mempelajari nada.

Dalam pengenalan suara, terdapat dua proses yaitu proses ekstraksi dan identifikasi. Hal ini juga terjadi dalam identifikasi nada yang dihasilkan oleh suara alat musik. Beberapa metode yang sering digunakan dalam ekstraksi sinyal suara adalah Mel-

Frequency Cepstrum Coefficients (MFCC) dan Linear Predictive Cepstral Coefficients (LPCC), sedangkan metode yang sering digunakan dalam penelitian terdahulu untuk identifikasi adalah MFCC.

MFCC merupakan metode untuk ekstraksi. MFCC telah digunakan untuk ekstraksi pada pengenalan suara burung sehingga memberikan akurasi sebesar 90% [1], mengoperasikan kursor dengan akurasi 88,89% [2]. Penelitian lain membandingkan fitur Linear Predictive Cepstral Coefficients (LPCC) terhadap MFCC untuk ekstraksi fitur yang memperlihatkan LPCC memiliki tingkat akurasi lebih tinggi yaitu 94,23 %, sedangkan MFCC sebesar 89,14%. Namun demikian, pada kondisi bising sistem berbasis MFCC memberikan kinerja yang lebih baik dibandingkan dengan LPCC. Pada tingkat bising yang cukup tinggi sistem berbasis MFCC memiliki tingkat akurasi 97,03% [3].

Dalam identifikasi sinyal suara, beberapa penelitian menggunakan Hidden Markov Model (HMM) untuk pengontrol lampu otomatis dengan menggunakan lima *state* pada HMM mendapatkan akurasi sebesar 98,84% [4] dan pengenalan ucapan untuk perintah gerak robot mobil penjejak dengan hasil pengenalan ucapan sebesar 83,33% [5], pengenalan ucapan pada permainan teka-teki silang berbasis desktop dengan akurasi sebesar 97,65% [6] dan penerapan HMM dalam *clustering sequence* protein globin [7].

Penelitian ini telah membangun sistem identifikasi nada yang dihasilkan alat musik menggunakan MFCC untuk ekstraksi serta HMM dengan K-Means *clustering* sebagai metode identifikasi. Sinyal suara yang dihasilkan oleh alat musik akan dilewatkan praproses DC-Removal untuk mereduksi variabel DC. Sedangkan Pre-emphasize untuk mempertahankan frekuensi ciri pada sebuah *spectrum*. Hasil dari praproses diekstraksi menggunakan MFCC.

Luaran dari MFCC dilakukan klasifikasi tiap *frame* menggunakan K-Means sehingga menghasilkan deret observasi. Deret observasi akan diklasifikasi menggunakan HMM. dengan jumlah kelas sebanyak 12 nada, yaitu. “C”, “C#”, “D”, “D#”, “E”, “F”, “F#”, “G”, “G#”, “A”, “A#” dan

“B” yang merupakan nada *major*, nada *major* digunakan karena nada ini adalah nada dasar dalam seni musik. Setiap nada akan direkam dengan alat musik Gitar, Bass, Biola, Piano, Ukulele.

II. METODE

A. Akuisisi Data

Penelitian terdahulu melakukan identifikasi suara untuk pengontrol lampu menggunakan MFCC dan HMM, menghasilkan akurasi sebesar 98,84%. Perolehan data direkam melalui *microphone* dengan frekuensi *sampling* sebesar 8000Hz menggunakan *channel* mono dan setiap file disimpan dengan format .wav [4]. Penelitian lain menggunakan frekuensi *sampling* sebesar 8000Hz dengan format .wav, *channel* mono, durasi dua detik dan resolusi 16 bit [8].

Setiap data diperoleh dengan melakukan perekaman suara menggunakan *microphone* dengan *channel* mono dari lima alat musik instrumen, yaitu Gitar, Bass, Biola, Piano dan Ukulele. Masing-masing alat musik membunyikan 12 nada dengan 10 kali perulangan untuk setiap nada dengan lama perekaman sebesar 2 detik, sehingga didapatkan 600 data perekaman yang digunakan sebagai data latih. Pada saat perekaman, nada yang digunakan adalah nada *major* “C”, “C#”, “D”, “D#”, “E”, “F”, “F#”, “G”, “G#”, “A”, “A#”, “B”, dengan oktaf yang tidak diatur. Hal tersebut dilakukan untuk memperoleh data latih yang baik. Setiap perekaman akan menghasilkan data *sampling* sebanyak 16000 dengan frekuensi *sampling* yang digunakan adalah 8000Hz.

B. Perancangan Sistem Identifikasi Nada

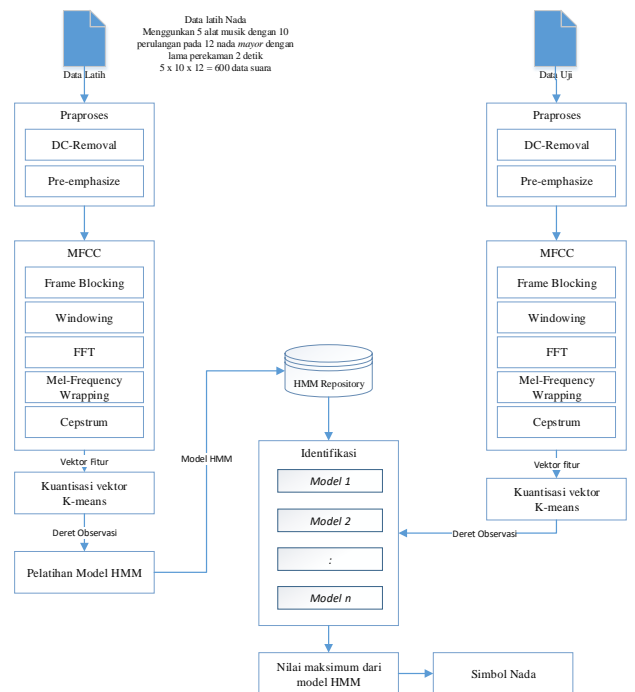
Pada penelitian ini masukan untuk perangkat lunak adalah data perekaman sinyal suara dari lima alat musik Gitar, Bass, Biola, Piano dan Ukulele. Perekaman dilakukan melalui *microphone* dengan frekuensi *sampling* 8000Hz, resolusi 16bit serta menggunakan *channel* mono. Setiap data masukan dilakukan praproses yaitu DC Removal dan Pre-Emphasize. DC Removal dilakukan untuk menghilangkan koponen DC sehingga diperoleh normalisasi dari data masukan suara, sedangkan Pre-Emphasize dilakukan untuk mempertahankan frekuensi ciri, sehingga dihasilkan data masukan suara dengan *noise* yang rendah.

Tahap selanjutnya setelah dilakukan praproses DC Removal dan Pre-Emphasize, dilakukan tahap ekstraksi fitur pada data perekaman sinyal suara yang dihasilkan oleh alat musik. Tahap ekstraksi dalam penelitian ini menggunakan MFCC. MFCC memiliki beberapa tahap dalam prosesnya yaitu Frame Blocking, Windowing, Fast Fourier Transform(FFT), Mel-Frequency Wrapping dan Cepstrum. Frame Blocking digunakan untuk segmentasi sinyal suara kedalam beberapa *frame* dengan panjang *sampling* untuk setiap *frame* 0,02 detik dan *overlap* sebesar 50%, sehingga didapatkan 199 *frame* pada setiap data perekaman. Windowing digunakan untuk meminimalisir diskontinuitas setiap *frame* pada tahap Frame Blocking. Pada tahap Windowing metode yang digunakan

adalah Hamming Window. Hamming Window dipilih karena menghasilkan *noise* yang tidak terlalu tinggi pada sinyal suara. FFT digunakan untuk mengkonversi sinyal suara dari domain waktu kedalam domain frekuensi. Mel-Frequency Wrapping pada tahap ini metode yang digunakan adalah Mel-Filterbank dengan jumlah *filter* yang digunakan adalah 32 *filter*. Cepstrum dilakukan pada sinyal suara untuk mendapatkan informasi berupa koefisien dengan koefisien yang digunakan adalah 13 koefisien untuk setiap *frame*.

Sebelum masuk pada tahap pelatihan dan pengujian menggunakan HMM, hasil ekstraksi akan dilakukan kuantisasi vektor menggunakan K-Means. Setiap data suara tersusun dari sejumlah vektor koefisien yang akan diterjemahkan kedalam sebuah deret observasi menggunakan algoritma K-Means. Hasil dari K-Means digunakan sebagai masukan untuk HMM.

Proses pelatihan dan pengujian dilakukan menggunakan HMM. Pelatihan dilakukan terhadap data dari hasil kuantisasi vektor berupa deret observasi menggunakan algoritma K-Means, di mana panjang dari deret observasi sama dengan jumlah *frame* yang ada pada proses ekstraksi dan deret observasi tersebut akan dimodelkan untuk setiap kelasnya. Kelas yang digunakan adalah sebanyak 12 kelas, yaitu “C”, “C#”, “D”, “D#”, “E”, “F”, “F#”, “G”, “G#”, “A”, “A#”, “B”. Proses pengujian dilakukan terhadap data yang telah dilatih dan data baru. Tahap praproses, ekstraksi, kuantisasi vektor dan identifikasi dapat dilihat pada Gambar 1.



Gambar 1. Perancangan Sistem Identifikasi Nada alat musik instrumen

C. Praproses

DC Removal dilakukan untuk menghilangkan komponen DC pada sinyal suara dengan cara, mengurangi tiap data sampel suara dengan nilai rata-rata data sampel suara. Penelitian terdahulu mengenai pengenalan tutur vokal bahasa

Indonesia menggunakan metode Discrete Wavelet Transform (DWT) dan Dynamic Time Warping (DTW) setelah dilakukan DC Removal mendapatkan akurasi untuk *tester trainer* sebesar 80% dan 87% untuk *tester* acak. DC Removal Dilakukan menggunakan Persamaan 1.

$$D[i] = s[i] - \frac{\sum_{i=1}^n s[i]}{n} \quad (1)$$

Dimana,

$D[i]$ = adalah hasil sinyal ke- i setelah dilakukan DC Removal,
 $s[i]$ = adalah sinyal awal ke- i ,
 n = adalah jumlah sample $n > 0$.

Pre-Emphasize dilakukan untuk mengurangi sinyal suara dari gangguan *noise* dengan mempertahankan frekuensi ciri pada sinyal suara yang umumnya tereliminasi pada saat proses perekaman, penelitian terdahulu melakukan identifikasi suara dengan nilai α 0,97 pada proses Pre-Emphasize [8]. Perhitungan Pre-Emphasize menggunakan Persamaan 2.

$$y[n] = s[n] - \alpha s[n - 1] \quad (2)$$

Dimana,

$y[n]$ = adalah sinyal hasil Pre-Emphasize,
 $s[n]$ = adalah data sampel ke- n .

Data sinyal suara hasil Pre-Emphasize adalah data sinyal suara sebelum proses Pre-Emphasize ditambah dengan data hasil Pre-Emphasize dengan Persamaan 2, sehingga hasil sinyal suara setelah Pre-Emphasize menggunakan Persamaan 3.

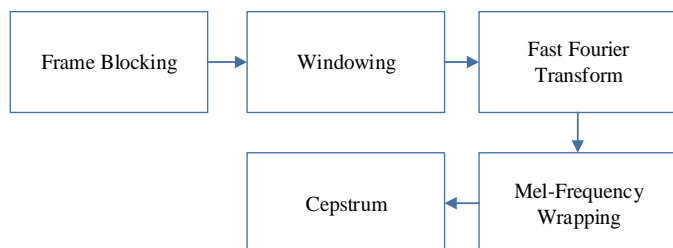
$$X[n] = X[n] + y[n] \quad (3)$$

Dimana,

$X[n]$ = Sinyal suara sebelum Pre-Emphasize,
 $y[n]$ = Sinyal hasil Pre-Emphasize.

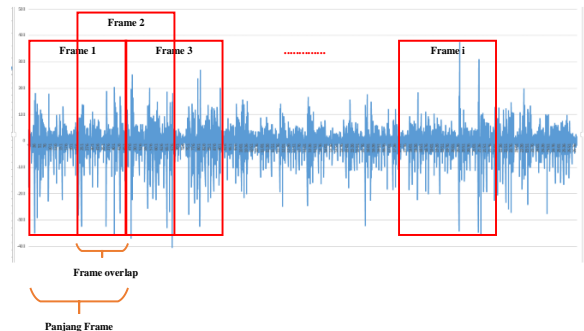
D. Mel-Frequency Cepstrum Coefficients

Dalam penggunaan MFCC dan dikombinasikan dengan perhitungan yang akurat akan memberikan tingkat keberhasilan yang tinggi, hal tersebut menjadikan Metode MFCC sebagai pilihan standar untuk ekstraksi fitur dalam beberapa penelitian pengenalan suara [9]. Penelitian terdahulu menggunakan MFCC sebagai ekstraksi [10][11][12]. Dalam Mel-Frequency Cepstrum Coefficients terdapat lima tahap yaitu, Frame Blocking, Windowing, FFT, Mel-Frequency Wrapping dan Cepstrum ditunjukkan pada Gambar 2.



Gambar 2. Proses MFCC

Pada tahap Frame Blocking sinyal suara dibagi ke dalam *frame-frame* dengan waktu tertentu yang lebih singkat. Proses Frame Blocking diawali dengan menentukan panjang frame yaitu sebesar 20 milidetik dan besaran *overlap* 50% dan menghasilkan 199 *frame* proses Frame blocking ditunjukkan pada Gambar 3.



Gambar 3. Frame Blocking

Windowing adalah proses untuk meminimalisir diskontinuitas pada permulaan dan akhir *frame* menggunakan Persamaan 4

$$y(n) = x(n)w(n), 0 \leq n \leq N - 1 \quad (4)$$

Dimana,

$x(n)$ = nilai sampel ke- i ,
 $w(n)$ = nilai window ke- n ,
 N = jumlah sampel dalam *frame*.

Metode *windowing* yang digunakan yaitu metode Hamming Window. Penggunaan Hamming window dapat meminimalkan efek kebocoran spektral yang diakibatkan dari proses Frame Blockin [13] yang ditunjukkan pada Persamaan 5.

$$w(n) = 0,5 - 0,46 \cos \left[\frac{2\pi n}{N-1} \right], 0 \leq n \leq N - 1 \quad (5)$$

Dimana,

$w(n)$ = nilai window ke- n ,
 N = jumlah sampel dalam *frame*.

Fast Fourier Transform merupakan proses untuk mengonversi setiap *frame* N sampel dari domain waktu ke domain frekuensi. Fast Fourier Transform merupakan *fast algorithm* dalam implementasi Discrete Fourier Transform (DFT) yang dioperasikan pada sebuah sinyal waktu – diskrit yang terdiri dari N sampel dengan Persamaan 6.

$$x(m) = \sum_{n=0}^{N-1} x(n) \cdot \left[\cos \left(\frac{2\pi mn}{N} \right) - j \sin \left(\frac{2\pi mn}{N} \right) \right] \quad (6)$$

Mel-Frequency Wrapping melakukan *filter* pada sinyal untuk setiap *frame* menggunakan Mel-Filterbank yang terdiri dari *triangular window* sebanyak N . *triangular window* yang digunakan yaitu sebanyak 32 buah dengan 34 titik dan untuk batas bawah dan batas atas nilai Mel ditentukan antara 0Hz dan

4000Hz yang dikonversi ke dalam nilai Mel menggunakan Persamaan 7.

$$mel(f) = 1127 * \ln\left(1 + \left(\frac{f}{700}\right)\right) \quad (7)$$

Dimana,

$mel(f)$ = Nilai *mel*, konversi dari nilai frekuensi,
 f = Nilai frekuensi.

Dalam Mel-Frequency Wrapping untuk mendapatkan nilai frekuensi dari mel menggunakan Persamaan 8.

$$mel^{-1}(f) = 700 \left(\exp\left(\frac{mel(f)}{1127}\right) - 1 \right) \quad (8)$$

Dimana,

$mel(f)$ = Nilai *mel*, konversi dari nilai frekuensi,
 $mel^{-1}(f)$ = Nilai frekuensi, konversi dari nilai mel,
 f = Nilai frekuensi.

Dalam pembuatan *filterbank*, nilai frekuensi (f) dikonversikan kedalam nilai sampel FFT terdekat menggunakan Persamaan 9.

$$f[m] = \left(\frac{N}{f_s}\right) mel^{-1}\left(mel(f_i) + m \frac{mel(f_n) - mel(f_i)}{M+1}\right) \quad (9)$$

Dimana,

$f[m]$ = nilai FFT terdekat,
 N = jumlah FFT tiap *frame*,
 f_n = frekuensi batas atas,
 f_i = frekuensi batas bawah,
 M = jumlah *filter*.

Dalam menghitung nilai dari *filterbank* menggunakan Persamaan 10.

$$H_{m[k]} = \begin{cases} 0 & k < f[m-1] \\ \frac{k-f[m-1]}{f[m]-f[m-1]} & f[m-1] \leq k \leq f[m] \\ \frac{f[m+1]-k}{f[m+1]-f[m]} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1] \end{cases} \quad (10)$$

Dimana,

$H_{m[k]}$ = nilai *filterbank*,
 m = indeks filter,
 k = indeks input sampel FFT,
 $f[m]$ = nilai FFT bin ke- m .

Tahap selanjutnya adalah tahap *filtering*, *filtering* digunakan untuk mendapatkan nilai *log energy* pada setiap *filter* dengan menggunakan Persamaan 11.

$$s[m] = \ln[\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k]], \quad 1 \leq m \leq M \quad (11)$$

Dimana,

$s[m]$ = nilai *log energy*,
 $X_a[k]$ = nilai FFT,
 $H_m[k]$ = nilai *filter*,
 N = jumlah nilai FFT,
 m = indeks *filter*,

k = indeks input sampel FFT.

Pada tahap Cepstrum nilai mel akan kembali dikonversikan kembali kedalam domain waktu, dan hasil dari proses Cepstrum ini disebut dengan Mel-Frequency Cepstrum Coefficients. Pada proses konversi menjadi Cepstrum teknik yang digunakan adalah Discrete Cosine Transform (DCT) dengan menggunakan Persamaan 12.

$$c_n = \sum_{k=1}^K (S[k]) \cos\left[\frac{\pi n(m+\frac{1}{2})}{K}\right], \quad n = 1, 2, \dots, K \quad (12)$$

Dimana,

c_n = Nilai koefisien Cepstrum ke- n ,
 $S[k]$ = Nilai *log energy*,
 n = indeks koefisien Cepstrum,
 k = indeks *filter*,
 K = jumlah *filter*.

E. K-Means

K-Means adalah suatu metode *clustering*, metode ini membagi data setiap *frame* dari ekstraksi MFCC menjadi *cluster-cluster* dengan *centroid* masing-masing *frame* yang nantinya menghasilkan sebuah deret observasi $O = \{O_{t_1}, O_{t_2}, O_{t_3}, \dots, O_{t_n}\}$ sebagai masukan untuk HMM. penelitian terdahulu menggunakan K-Means sebagai *classifire* [14][15]. Tahap pertama dalam K-Means yaitu menentukan *cluster* secara sembarang. Kemudian hitung jarak terdekat dengan *cluster* menggunakan Persamaan 13.

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (X_{ij} - C_{ij})^2} \quad (13)$$

Dimana,

C_{ij} = *centroid*.

Lalu tahap selanjutnya menentukan nilai *centroid* yang baru. Dalam menentukan *centroid* baru dapat dicari dengan cara menghitung nilai rata-rata dari data yang menjadi anggota untuk *cluster* tersebut menggunakan Persamaan 14.

$$C_{ij} = \frac{\sum_{i=1}^p X_{ij}}{p} \quad (14)$$

Dimana $X_{ij} \in \text{cluster ke-}k$,

P = banyaknya anggota *cluster* ke- k .

F. Hidden Markov model

Hidden Markov Model (HMM) merupakan proses stokastik ganda, yaitu suatu matriks tak negatif yang jumlah masukan pada tiap kolomnya sama dengan 1 dan salah satu prosesnya tidak dapat diobservasi atau tersembunyi. Suatu proses yang tersembunyi tersebut hanya dapat diobservasi menggunakan suatu proses yang dapat diobservasi. Maka jika terdapat $X = \{X_1, X_2, \dots\}$ adalah sebuah proses Markov, dan $O = \{O_1, O_2, \dots\}$ adalah sebuah fungsi dari X , maka X adalah sebuah HMM yang dapat diobservasi melalui O , atau dapat ditulis $O = f(X)$ untuk suatu fungsi f . Parameter X menyatakan state process yang

tersembunyi (hidden), sedangkan O menyatakan observation process yang dapat diobservasi, HMM memiliki 5 komponen yaitu :

- **Hidden State (N)**
Hidden state merupakan state yang tersembunyi dan tidak dapat diamati.
- **Observed State (M)**
Banyaknya elemen keadaan yang terobservasi yang dinotasikan dengan M .
- **Matriks Peluang Transisi (A)**
Matriks peluang transisi $A = \{a_{ij}\}$ di mana a_{ij} merupakan peluang transisi bersyarat dari keadaan $t + 1$, jika diketahui keadaan X pada saat t , atau $a_{ij} = P(X_{t+1} = j | X_t = i)$ di mana $1 \leq i, j \leq N$. Karena itu A berukuran $N \times N$.
- **Matriks Peluang Emisi (B)**
Matriks peluang emisi adalah peluang perpindahan state pada saat t , pada keadaan i , yang dinotasikan dengan $B = \{b_i(k)\}$, di mana $b_i^k = b_i(k) = P(O_t = k | X_t = i)$, $1 \leq i \leq N$, $1 \leq k \leq M$. k adalah observasi pada waktu t bernilai k .
- **Peluang Keadaan Awal (π)**
Peluang keadaan awal di simbolkan dengan π dimana, $\pi(i) = P(X_1 = i)$, $1 \leq i \leq N$.

Berdasarkan kelima komponen tersebut, maka untuk merancang HMM dibutuhkan 5 komponen yaitu (N, M, A, B dan π). Namun terdapat tiga parameter yang menotasikan HMM yaitu A, B dan π . Sehingga HMM dinotasikan dengan $\lambda = (A, B, \pi)$.

Penelitian terdahulu menggunakan HMM sebagai sistem identifikasi pembicara berbasis power spektrum dengan perbandingan jumlah hidden state 3 memiliki akurasi 99,3%, 4 hidden state memiliki akurasi 99%, 5 hidden state memiliki akurasi 99,5%, 6 hidden state memiliki akurasi 100% dan 7 hidden state memiliki akurasi 99% [16]. Penelitian lain menggunakan HMM untuk named entity recognition pada teks tanaman obat [17], pengenalan pola suara berbasis fitur suara dengan akurasi sebesar 98.75% [18], pengenalan angka terisolasi [19] dan pengenalan ucapan arab menghasilkan akurasi sebesar 93.43% [20].

III. HASIL DAN DISKUSI

Penelitian ini menggunakan data latihan dengan jumlah 600 data perekaman yang diperoleh dari 5 alat musik instrumen, yaitu Gitar, Bass, Biola, Piano dan Ukulele. Setiap alat musik membunyikan 12 nada mayor yang telah ditentukan dengan setiap nada dilakukan perulangan sebanyak 10 kali. Pengujian dilakukan dengan menggunakan data perekaman yang telah dilatih dan data perekaman baru. Data baru didapatkan dari perekaman alat musik Gitar, Bass, Biola, Piano dan Ukulele yang membunyikan 12 nada dengan 10 kali perulangan untuk setiap nadanya.

Kuantisasi vektor terhadap data hasil ekstraksi MFCC menggunakan K-Means dengan 14 cluster dan maksimal epoch sebanyak 1000.

Uji optimalisasi parameter HMM dilakukan terhadap data latih dan data baru. Dengan jumlah hidden state 3 dan 6 dengan jumlah epoch 50, 100, 500, dan 1000. Hasil uji optimalisasi parameter HMM data latih dapat dilihat pada TABEL I dan data baru pada TABEL II.

TABEL I. HASIL UJI OPTIMALISASI DATA LATIH

No	Hidden State	Epoch	Rata-rata Akurasi(%)
1	3	50	78
2	3	100	78
3	3	500	78
4	3	1000	78
5	6	50	78
6	6	100	91
7	6	500	92
8	6	1000	92

TABEL II HASIL UJI OPTIMALISASI DATA BARU

No	Hidden State	Epoch	Rata-rata Akurasi(%)
1	3	50	57
2	3	100	57
3	3	500	57
4	3	1000	57
5	6	50	58
6	6	100	66
7	6	500	66
8	6	1000	66

Berdasarkan hasil uji parameter pelatihan HMM, parameter yang mendapatkan akurasi terbaik untuk uji parameter terhadap data latih yaitu dengan hidden state 6 epoch 500 dan hidden state 6 epoch 1000 dengan akurasi sebesar 92%. Sedangkan untuk akurasi terbaik untuk uji parameter terhadap data baru yaitu dengan hidden state 6 epoch 100, hidden state 6 epoch 500 dan hidden state 6 epoch 1000 mendapatkan rata-rata akurasi sebesar 66%. Akurasi didapat dari menghitung persentasi data yang dikenali dari semua data dengan menggunakan Persamaan 15.

$$akurasi = \frac{data\ yang\ dikenali \times 100}{jumlah\ data} \quad (15)$$

IV. KESIMPULAN

Penelitian ini telah menghasilkan sistem identifikasi nada dari sinyal suara alat musik instrumen menggunakan Mel – Frequency Cepstrum Coefficients dan Hidden Markov Model dan telah dilakukan proses pengujian. Proses pengujian dilakukan terhadap sistem dan pengujian akurasi sistem. Pengujian sistem dilakukan terhadap fungsionalitas sistem, sedangkan untuk pengujian akurasi sistem dilakukan terhadap data latih dan data baru serta parameter pelatihan HMM yang digunakan.

Parameter pelatihan HMM yang digunakan adalah jumlah *hidden state* sebanyak 6 dan jumlah *epoch* sebesar 1000. Dengan menggunakan parameter pelatihan tersebut didapatkan akurasi untuk data latih sebesar 92% dan data baru untuk alat musik Biola 74%, Ukulele 81%, Gitar 61%, Piano 82% dan Bass 30%.

Dari hasil pengujian yang telah dilakukan, maka didapatkan kesimpulan bahwa penggunaan jumlah dari *hidden state*, *epoch* dan kualitas perekaman yang tepat dapat mempengaruhi tingkat akurasi pada sistem identifikasi nada dari sinyal suara alat musik.

UCAPAN TERIMA KASIH

Terima kasih kepada Direktorat Jenderal Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi dan Pendidikan Tinggi atas pemberian dana Program Kreativitas Mahasiswa Bidang Penelitian Eksakta tahun 2018.

REFERENSI

- [1] F. Budiman, M. A. Nursyeha, M. Rivai, and Suwito, "Frequency Cepstrum Coefficient Dan Jaringan Syaraf Tiruan Pada Sistem Pengusir Hama Burung," vol. 5, no. 1, pp. 65–72, 2016.
- [2] A. Setiawan, A. Hidayatno, and R. R. Isnanto, "Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer," *Univ. Diponegoro Semarang*, vol. 13, no. 3, pp. 82–86, 2011.
- [3] U. Bhattacharjee, "A comparative study of LPCC and MFCC features for the recognition of Assamese phonemes," *Int. J. Eng. Res. Technol.*, vol. 2, no. 3, pp. 1–6, 2013.
- [4] A. K. Munggaran, E. C. Djamal, and R. Yuniarti, "Identifikasi Suara Pengontrol Lampu Menggunakan Mel-Frequency Cepstral Coefficients dan Hidden Markov Model," pp. 17–22, 2017.
- [5] Fitrilina, R. Kurnia, and S. Aulia, "Identifikasi, Pengenalan Ucapan Metoda Mfcc-Hmm Untuk Perintah Gerak Robot Mobil Penjejak," *J. Nas. Tek. Elektro*, vol. 2, no. 1, pp. 31–40, 2013.
- [6] M. T. S. Jaya, D. Puspitaningrum, and B. Susilo, "Penerapan Speech Recognition Pada Permainan Teka-Teki Silang Menggunakan Metode Hidden Markov Model (Hmm) Berbasis Desktop," *J. Rekursif*, vol. 4, no. 1, pp. 119–129, 2016.
- [7] S. Mulyana, Afiahayati, and W. A. Surya, "Penerapan Hidden Markov Model Dalam Clustering Sequence Protein Globin," *Seminar*, no. 2004, pp. 117–122, 2008.
- [8] A. D. Andriana, "Perangkat Lunak Untuk Membuka Aplikasi Pada Komputer Dengan Perintah Suara Menggunakan Metode Mel Frequency Cepstrum Coefficients," *Ilm. Komput. dan Inform.*, vol. 2, no. 1, pp. 21–26, 2013.
- [9] T. D. Ganchev, *Speaker Recognition*, no. November. 2005.
- [10] Z. Effendi, T. Erlina, and R. Aisuwarya, "Pengenalan Suara Menggunakan Metode MFCC (Mel Frequency Cepstrum Coefficients) dan DTW (Dynamic Time Warping) untuk Sistem Penguncian Pintu ISBN : 979-26-0280-1 ISBN : 979-26-0280-1," pp. 239–243, 2015.
- [11] A. Maesa, F. Garzia, M. Scarpiniti, and R. Cusani, "Text Independent Automatic Speaker Recognition System Using Mel-Frequency Cepstrum Coefficient and Gaussian Mixture Models," vol. 2012, no. October, pp. 335–340, 2012.
- [12] Ranny, "Voice Recognition using k Nearest Neighbor and Double Distance Method," 2016.
- [13] M. W. . Sanjaya and Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (Mfcc) Dan Adaptive Neuro-Fuzzy Inferense System (Anfis) Sebagai Kontrol Lampu Otomatis," *Al-HAZEN J. Phys.*, vol. 1, no. 1, pp. 1–19, 2014.
- [14] A. Rizal and V. Suryani, "Pengenalan Signal Ekg Menggunakan Dekomposisi Paket Wavelet Dan K-Means-Clustering," *Semin. Nas. Apl. Teknol. Inf.*, vol. 2008, no. Snati, pp. 1907–5022, 2008.
- [15] A. Alsayat, "Efficient Genetic K-Means Clustering for Health Care Knowledge Discovery," 2016.
- [16] A. Buono, W. Jatmiko, and B. Kusumoputro, "Sistem Identifikasi Pembicara Berbasis Power Spektrum Menggunakan Hidden Markov Model," vol. 77, pp. 257–286, 2010.
- [17] A. Setiyoaji, L. Muflikhah, and M. A. Fauzi, "Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat," vol. 1, no. 12, pp. 1858–1864, 2017.
- [18] X. Li, M. Yao, and J. Yang, "Speech Recognition Approach Based on Speech Feature Clustering and HMM," *J. Comput.*, vol. 7, no. 9, pp. 2269–2275, 2012.
- [19] S. J. Lf and A. P. Suara, "Pengenalan Angka Terisolasi dengan Menggunakan Pemodelan HMM Melalui Ekstraksi Feature Mel-Cepstrum Filter Bank," pp. 1–7, 2011.
- [20] M. Abushariah, R. N. Aionon, R. Zainuddin, M. Elshafei, and O. O. Khalifa, "Natural speaker-independent Arabic speech recognition system based on Hidden Markov Models using Sphinx tools," *Int. Conf. Comput. Commun. Eng.*, no. May, pp. 1–6, 2010.