

**SISTEM PENGENALAN *CHORD* PADA *FILE* MUSIK DIGITAL
DENGAN MENGGUNAKAN *PITCH CLASS PROFILES* DAN
*HIDDEN MARKOV MODEL***

Ivanna K. Timotius, Adhi Prayogo

Fakultas Teknik Elektronika dan Komputer, Program Studi Teknik Elektro,

Universitas Kristen Satya Wacana, Salatiga, Indonesia

ivanna_timotius@yahoo.com, adhi.prayogo@gmail.com

Intisari

Chord adalah salah satu elemen utama musik yang dapat memberikan gambaran sederhana dari suatu karya musik. Sebuah sistem yang dapat mengenali *chord* dari suatu rekaman musik secara otomatis akan sangat menarik dan bermanfaat. Tulisan ini mengimplementasikan sebuah algoritma pengenalan *chord* otomatis dengan menggunakan *pitch class profiles* dan *hidden Markov model*. Algoritma yang diimplementasikan menunjukkan tingkat akurasi pengenalan sebesar 100% untuk file lagu progresi *chord* murni dari MIDI, 97,97% untuk file lagu progresi *chord* murni dari rekaman gitar, 85,35% untuk file lagu kompleks dari MIDI, dan 59,80% untuk file lagu kompleks yang mengandung vokal manusia.

Kata kunci: *chord*, *pitch class profiles*, *hidden Markov model*.

1. Pendahuluan

Sistem pendengaran manusia mempunyai kemampuan untuk menyarikan data yang kaya dari sinyal audio yang kompleks. Misalnya manusia dapat memahami percakapan dan menikmati musik. Musik maupun percakapan termasuk sinyal audio yang kompleks. Salah satu elemen utama musik adalah *chord*. Rangkaian *chord* dapat memberikan gambaran sederhana dari sebuah karya musik.

Transkripsi musik didefinisikan sebagai kegiatan mendengarkan sebuah karya musik dan menulis notasi musik untuk setiap nada yang menyusun karya musik tersebut.

Transkripsi musik dapat juga diartikan sebagai kegiatan mentransformasi unsur – unsur dari sebuah karya musik ke dalam representasi simboliknya. Transkripsi *chord* adalah bagian dari transkripsi musik yang bertujuan mendapatkan representasi simbolik dari unsur harmoni sebuah karya musik, yaitu nama - nama *chord* dari karya musik tersebut. Transkripsi *chord* ini penting dan bermanfaat misalnya bagi pemain musik yang ingin mempelajari dan memainkan kembali lagu dari suatu rekaman musik.

Metode yang digunakan hingga kini dalam melakukan transkripsi *chord* adalah dengan transkripsi secara manual. Caranya adalah seseorang mendengarkan suatu rekaman musik lalu menuliskan *chord*-nya dengan mengandalkan kepekaan pendengaran dan perasaannya. Tetapi orang yang dapat melakukan ini adalah orang – orang tertentu dengan kepekaan telinga yang baik dan memiliki latar belakang musik yang kuat serta sudah terlatih secara khusus dalam jangka waktu lama. Karena itu sebuah sistem yang dapat melakukan transkripsi *chord* secara otomatis akan sangat bermanfaat untuk membantu orang – orang yang baru belajar musik atau pemusik pemula untuk mendapatkan label *chord* yang benar dari suatu rekaman musik yang dipunyai.

Sistem yang dikembangkan menggunakan *Pitch Class Profiles* (PCP) sebagai metode ekstraksi fitur. PCP dipilih karena dapat menampilkan intensitas nada – nada musikal dari suatu sinyal dengan baik. Hal ini penting karena yang membedakan suatu *chord* dari *chord* yang lain adalah nada – nada penyusunnya. Selain itu, PCP memetakan nada – nada yang sama dari oktaf yang berlainan ke dalam satu kelas nada, hal ini juga akan sangat berguna karena *chord* akan tetap sama jika nada – nada penyusunnya sama, tidak bergantung pada urutan penyusunannya atau jangkauan oktafnya.

Chord dalam sebuah rekaman musik berubah mengikuti waktu, sebagaimana halnya kata – kata dalam rekaman percakapan berubah mengikuti waktu, karena itu banyak kerangka kerja dalam pengenalan wicara yang dapat digunakan dengan modifikasi [1]. Sistem yang dikembangkan menggunakan *Hidden Markov Models* (HMM) sebagai klasifikator yang populer digunakan dalam pengenalan wicara [2].

HMM dapat dipandang sebagai pernyataan *state machine* yang *state*-nya berpindah – pindah setiap waktu tertentu. Perpindahan antar *state* dalam sebuah HMM berdasarkan probabilitas tertentu. Hal ini juga bersesuaian dengan sifat *chord* dalam

SISTEM PENGENALAN CHORD PADA FILE MUSIK DIGITAL DENGAN MENGGUNAKAN PITCH CLASS PROFILES DAN HIDDEN MARKOV MODEL

Ivanna K. Timotius, Adhi Prayogo

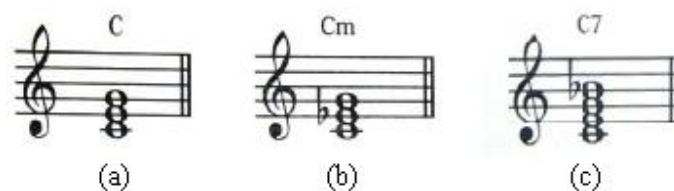
sebuah lagu yang berpindah – pindah setiap waktu tertentu berdasarkan probabilitas tertentu. Suatu *chord* dalam sebuah lagu akan mempunyai kecenderungan lebih kuat untuk berpindah ke suatu *chord* tertentu daripada ke *chord* lainnya.

Terdapat beberapa kesulitan yang dihadapi suatu sistem pengenalan *chord* otomatis di antaranya adalah banyaknya jenis alat musik yang digunakan dengan warna suara yang berbeda-beda, adanya alat musik perkusi, adanya *chord* yang nada-nadanya tidak dimainkan secara simultan (*broken chord* atau *arpeggio*), dan adanya melodi lagu.

2. Chord

Chord merupakan sekumpulan nada dengan pola tertentu. Sinyal dari sebuah nada yang berbeda memiliki frekuensi yang berbeda. Karena itu, sinyal dari dua buah *chord* yang berbeda, yang disusun oleh nada yang berbeda, akan memiliki karakteristik frekuensi yang berbeda pula.

Chord dalam musik adalah tiga atau lebih nada – nada yang dibunyikan bersama – sama [3]. *Chord* diklasifikasikan berdasarkan interval di antara nada – nada pembentuknya. Interval dalam musik berarti jarak di antara dua nada musikal. Jenis *chord* yang paling umum dan paling sederhana adalah triad, yaitu yang terdiri dari tiga buah nada. Dua jenis triad yang paling umum adalah major triad dan minor triad. Bila sebuah interval ditambahkan lagi pada sebuah triad, maka akan membentuk *seventh chord* yang terdiri dari empat nada. *Seventh chord* yang paling umum adalah *dominant seventh chord*. Contoh major triad, minor triad, *dominant seventh chord* dengan *root* C yang dituliskan dalam notasi balok ditunjukkan pada Gambar 1.



Gambar 1. Notasi Balok untuk (a) C Major Triad (b) C Minor Triad (c) C Dominant Seventh.

3. Pitch Class Profiles (PCP)

Metode PCP bertujuan untuk membentuk sebuah vektor PCP yang memiliki beberapa elemen pada setiap region waktu tertentu [4]. Tiap elemen pada vektor ini mewakili intensitas sinyal dari region-region frekuensi yang telah ditetapkan. Apabila tiap nada pada tangga nada kromatis direpresentasikan oleh sebuah elemen di vektor PCP, maka vektor PCP ini akan memiliki 12 elemen.

Metode PCP dimulai dengan membagi *file* lagu menjadi beberapa *frame* dengan ukuran N . Data-data pada tiap *frame* tersebut $x[n]$ ditransformasikan ke ranah frekuensi dengan menggunakan *Short Time Fourier Transform* (STFT) dengan menggunakan suatu jendela $w[n]$:

$$X_{STFT}[k, n] = \sum_{m=0}^{N-1} x[n-m] \cdot w[m] \cdot e^{-j2\pi km/N} \quad 0 \leq k \leq N-1 \quad (1)$$

Jenis jendela yang digunakan pada sistem ini adalah jendela Hanning. Jenis jendela ini dipilih karena mempunyai resolusi amplitudo yang bagus dan resolusi frekuensi yang cukup baik. Jendela Hanning dengan orde N diberikan oleh persamaan berikut:

$$w[m] = 0.5 \left[1 - \cos \left(\frac{2\pi m}{N-1} \right) \right] \quad 0 \leq m \leq N-1 \quad (2)$$

Setiap data pada sinyal hasil STFT terjendela kemudian dipetakan ke dalam kelas nada (*pitch class*). Kelas nada dari masing-masing data dipetakan dengan menggunakan persamaan:

$$p[k] = \left[N_c \cdot \log_2 \left(\frac{k}{N} \cdot \frac{f_{sr}}{f_{ref}} \right) \right] \bmod N_c \quad (3)$$

dengan $p[k] = 0, 1, 2, \dots, N_c - 1$ merupakan kelas nada dari sampel k pada sinyal hasil STFT terjendela, N_c merupakan jumlah kelas nada, N merupakan resolusi dari STFT yang sama dengan ukuran *frame*, f_{sr} merupakan frekuensi cuplik dari sinyal masukan, dan f_{ref} merupakan frekuensi referensi, misalnya untuk nada A $f_{ref} = 440$ Hz.

Nilai elemen-elemen pada vektor PCP $x_{pcp}[i]$, $0 \leq i \leq N_C - 1$, didapatkan dengan merata-ratakan magnitudo dari sinyal di ranah frekuensi $X[k]$ yang memiliki kelas nada yang sama:

$$x_{pcp}[i] = \frac{1}{\#(k \in (p[k] = i))} \sum_{p[k]=i} |X[k]|^2 \quad (4)$$

4. Hidden Markov Model (HMM)

HMM merupakan sebuah sistem pemodelan yang bertujuan untuk mengenali suatu pola dengan menggunakan proses pembelajaran terlebih dahulu. Sebuah HMM adalah sebuah proses acak ganda (*doubly stochastic process*) [5][6] dengan satu proses acak pokok yang tidak dapat diobservasi (*hidden*), tetapi hanya dapat diobservasi melalui himpunan proses acak lainnya yang menghasilkan urutan simbol yang terobservasi. HMM juga dapat dipandang sebagai sebuah automaton acak terbatas dimana setiap *state*-nya menghasilkan suatu observasi. Transisi antar *state* mematuhi properti Markov, yaitu perpindahan antar *state* pada HMM orde n hanya bergantung pada n *state* sebelumnya.

Salah satu tipe HMM adalah *ergodic* HMM dimana setiap *state* dapat dicapai dari setiap *state* lainnya dalam sejumlah langkah tertentu. Pada penerapan konsep HMM untuk membangun aplikasi pengenalan *chord*, setiap *chord* diwakili oleh sebuah *state* dalam proses acak utama. Proses acak utama adalah proses transisi antar *chord*, yang mematuhi properti Markov. Proses acak utama ini tersembunyi. Pada setiap saat t , *state* aktif akan menghasilkan sebuah vektor PCP, vektor PCP inilah yang merupakan urutan simbol yang dapat diobservasi.

Sebuah HMM I terdiri dari parameter – parameter A , B , dan p . Pernyataan ini sering ditulis dengan

$$\lambda = (A, B, p) \quad (5)$$

Parameter A adalah matriks transisi dari *hidden status*, dengan $A = \{a_{ij}\}$ dan a_{ij} adalah peluang *next state* q_j jika *present state* q_i

$$a_{ij} = P[q_j|q_i] \quad (6)$$

q adalah elemen dari himpunan *hidden state* $Q = \{q_1, q_2, q_3, \dots, q_N\}$ dengan N adalah jumlah seluruh *hidden state*. Parameter B sering disebut dengan *confusion matrix* memuat distribusi probabilitas dari simbol keluaran yang terobservasi. Parameter B memberikan informasi peluang munculnya keluaran terobservasi tertentu pada saat *state* tertentu:

$$B = \{b_{ij}\} \quad (7)$$

$$b_{ij} = P[v_j | q_i] \quad (8)$$

b_{ij} adalah peluang munculnya v_j jika sedang berada pada *state* q_i . Parameter p adalah vektor probabilitas *state* awal. Jadi maksudnya adalah peluang *state* pertama adalah *state* tertentu:

$$p_i = P[S_1 = q_i] \quad (9),$$

S_1 merupakan *state* pertama. Dalam sistem pengenalan *chord*, p berarti peluang sebuah lagu dimulai dengan *chord* tertentu.

5. Implementasi dan Pengujian

File musik digital yang digunakan sebagai masukan sistem diubah menjadi *file* musik mono dalam format wav dengan frekuensi cuplik 11025 Hz dan kuantisasi 8 bit. Frekuensi cuplik sebesar 11025 Hz dipilih untuk mendapatkan lebar pita sinyal sekitar 5500 Hz yang sudah cukup untuk mendapatkan informasi harmoni dari nada – nada musik. Selain itu lebar pita sekitar 5500 Hz dapat menghilangkan pengaruh bunyi pada frekuensi tinggi seperti bunyi simbal. Penggunaan *file* masukan mono dengan frekuensi cuplik sebesar 11025 Hz, kuantisasi 8 bit dipilih untuk mempercepat proses komputasi.

Sistem didesain untuk dapat mengenali 3 jenis *chord families* yaitu *major chord*, *minor chord* dan *dominant seventh chord* pada 12 *roots* yaitu A, Bb, B, C, C#, D, Eb, E, F, F#, G, G#. *File* musik yang digunakan melatih sistem dan yang diuji berjumlah 36 *file* musik yang terdiri dari 4 kelompok, yaitu

1. 14 *file* musik berisi rekaman progresi *chord* murni yang direkam dari MIDI. *File* musik yang dimaksudkan pada kelompok ini adalah *file* lagu yang mengandung suara dari satu alat musik yang memainkan suatu urutan *chord* saja. *File* ini tidak mengandung melodi dan atau iringan alat musik perkusi. Urutan *chord* yang direkam

**SISTEM PENGENALAN CHORD PADA FILE MUSIK DIGITAL DENGAN
MENGUNAKAN PITCH CLASS PROFILES DAN HIDDEN MARKOV MODEL**

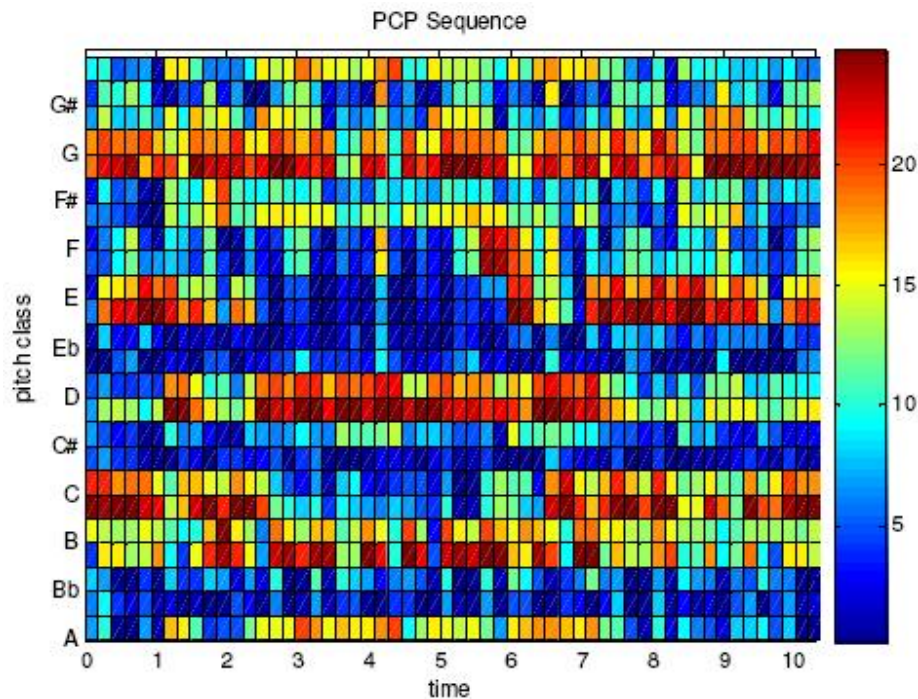
Ivanna K. Timotius, Adhi Prayogo

dipilih sedemikian sehingga dapat mewakili seluruh jenis *chord* yang ingin diklasifikasikan.

2. 14 *file* lagu berisi rekaman progresi *chord* murni yang direkam dari permainan gitar dengan progresi *chord* yang sama dengan kelompok sebelumnya.
3. 5 *file* lagu berisi lagu kompleks yang direkam dari MIDI. Lagu kompleks adalah lagu umum yang sudah mengandung melodi dan diiringi oleh beberapa alat musik, lengkap dengan alat musik perkusi. Lagu didapatkan dari *file – file* MIDI yang suaranya direkam menjadi *file* berformat *wav*.
4. 3 *file* lagu berisi lagu kompleks komersial yang mengandung suara vokal manusia dan diiringi oleh beberapa alat musik.

Pada sistem yang dikembangkan ini dipakai *frame* yang saling *overlap* dengan panjang *frame* $N = 4096$ dan data *ter-overlap* 50%. Dengan frekuensi cuplik dari sinyal masukan sebesar 11025 Hz, akan menghasilkan *frame* sinyal dengan kerapatan sekitar 5 *frame* per detik. Panjang *frame* ini dipilih karena lagu – lagu pada umumnya, *chord* baru akan berganti setelah beberapa ketukan. Jika sebuah sinyal lagu masukan mempunyai tempo 120 ketukan per menit dan ada bagian lagu yang *chord*-nya berganti setiap ketukan, maka dibutuhkan *frame* sinyal untuk setiap ketukan, yang dapat dicapai cukup dengan kerapatan $120/60 = 2$ *frame* per detik saja. Karena itu kerapatan 5 *frame* per detik yang diimplementasikan sudah memadai untuk menangani lagu – lagu pada umumnya.

Jumlah kelas nada N_C pada PCP adalah 24 ditentukan secara empirik sehingga menghasilkan akurasi yang terbesar. Gambar 2 menunjukkan hasil vektor PCP terhadap sumbu waktu dalam detik.



Gambar 2. Contoh Vektor PCP.

Proses ekstraksi fitur dari sinyal lagu masukan menghasilkan sekuen vektor PCP. Frame – frame PCP tersebut digunakan untuk membangun sebuah HMM yang memiliki satu *state* untuk setiap *chord*. Sistem yang diimplementasikan dirancang untuk dapat mengenali 36 jenis *chord*, oleh karena itu dibentuk sebuah HMM dengan 36 buah *state*.

Sesuai Persamaan 5, parameter A disimpan dalam sebuah matriks 36×36 yang tiap elemennya a_{ij} berisi peluang *state* berikut q_j jika *state* sekarang adalah q_i . Parameter B untuk tiap *state* dari 36 *state* diwakili oleh sebuah vektor yang berisi *mean* dari tiap *pitch class* untuk *state* tersebut, dan sebuah vektor yang berisi standar deviasinya. Parameter p disimpan dalam sebuah vektor dengan 36 buah elemen. Masing – masing elemen berisikan peluang *state* awal yang muncul pada proses pengenalan.

Hasil pengujian sistem yang telah dibangun ditunjukkan oleh Tabel 1. Hasil tersebut dibandingkan dengan kemampuan manusia dalam mengenali *chord*. Orang yang diuji adalah orang yang sudah terlatih musik selama jangka waktu yang bervariasi. Cara melakukan tes adalah dengan mendengarkan sejumlah lagu kepada para peserta, lalu peserta tes diminta langsung menuliskan *chord* dari lagu tersebut. Lagu yang diujikan berjumlah 8 buah dari jenis lagu kompleks yang terdiri dari 5 lagu dari hasil rekaman

**SISTEM PENGENALAN CHORD PADA FILE MUSIK DIGITAL DENGAN
MENGUNAKAN PITCH CLASS PROFILES DAN HIDDEN MARKOV MODEL**

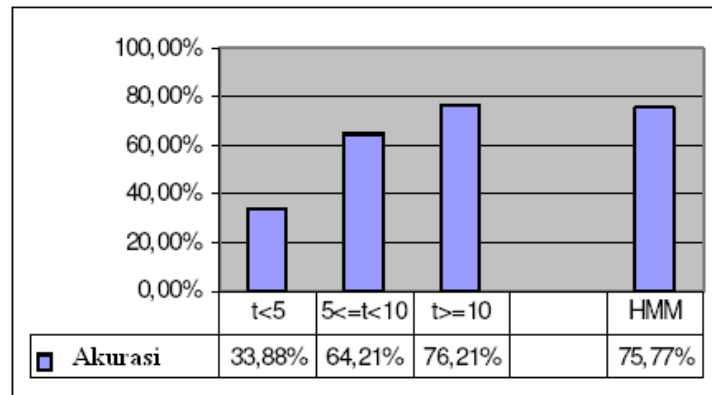
Ivanna K. Timotius, Adhi Prayogo

MIDI dan 3 lagu yang mengandung suara vokal manusia. Kepada peserta tes lagu diperdengarkan sebanyak dua kali, hal ini dilakukan untuk memberi kesempatan berpikir dan menulis. Nada dasar dari lagu yang diujikan juga diberitahukan terlebih dahulu kepada para peserta tes. Hal ini dilakukan karena orang yang mampu mengenali *chord* dari lagu tanpa diberitahukan nada dasar lagu terlebih dahulu adalah orang yang memiliki talenta istimewa dan sangat jarang ditemui. Tes diberikan kepada 10 orang yang terdiri dari 4 orang baru berlatih musik di bawah 5 tahun, 3 orang sudah berlatih musik antara 5 hingga kurang dari 10 tahun, dan 3 orang sudah berlatih musik selama 10 tahun atau lebih. Gambar 3 menunjukkan grafik perbandingan tingkat akurasi rata – rata pengenalan *chord* oleh manusia dari tiap kelompok orang dengan pengenalan *chord* otomatis yang berbasis HMM.

Dari grafik dapat dilihat bahwa tingkat akurasi sistem dalam mengenali *chord* kira – kira sebanding dengan orang yang berlatih musik selama 10 tahun. Dengan demikian sistem pengenalan *chord* otomatis yang telah direalisasikan dapat bermanfaat untuk membantu orang – orang yang belum begitu terlatih musik untuk mendapatkan *chord* yang benar dari suatu rekaman musik.

Tabel 1. Akurasi Sistem Pengenalan *Chord*.

Kelompok File Lagu	Akurasi (%)
Progresi <i>chord</i> murni dari MIDI	100
Progresi <i>chord</i> murni dari gitar	97,97%
Lagu kompleks dari MIDI	85,35%
Lagu kompleks dengan vokal manusia	59,80%



Gambar 3. Grafik perbandingan tingkat akurasi rata-rata pengenalan *chord* oleh manusia dan sistem pengenalan *chord* otomatis.

6. Kesimpulan

Chord dari sebuah file musik dapat dikenali secara otomatis dengan menggunakan ekstraksi fitur PCP dan HMM. Hasil pengujian terbaik memberikan tingkat akurasi 100% untuk file lagu progresi *chord* murni dari MIDI, 97,97% untuk file lagu progresi *chord* murni dari rekaman gitar, 85,35% untuk file lagu kompleks dari MIDI, dan 59,80% untuk file lagu kompleks yang mengandung vokal manusia.

Daftar Pustaka

1. A. Sheh dan D. P.W. Ellis, *Chord Segmentation and Recognition using EM-Trained Hidden Markov Models*, LabROSA, Dept. of Electrical Engineering, Columbia University.
2. B. Gold and N. Morgan, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, John Wiley & Sons, Inc., 1999.
3. Encarta Encyclopedia. Microsoft ® Encarta ® 2006. © 1993-2005, Microsoft Corporation.
4. G. Cabral, J. P. Briot dan F. Pachet, *Impact of Distance in Pitch Class Profile Computation*, Laboratoire d'Informatique de Paris, Université Pierre et Marie Curie, dan Sony Computer Science Lab Paris.
5. L. R. Rabiner dan B. H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, Januari 1986.

***SISTEM PENGENALAN CHORD PADA FILE MUSIK DIGITAL DENGAN
MENGUNAKAN PITCH CLASS PROFILES DAN HIDDEN MARKOV MODEL***
Ivanna K. Timotius, Adhi Prayogo

6. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceeding of IEEE*, vol. 77, no. 2, pp. 257-286 , Februari 1989.