

SEGMENTASI DAN KLASIFIKASI SPEECH / MUSIC MENGGUNAKAN ALGORITMA SEGMENTAL CDP

Allan Nafari¹, Bambang Hidayat², Gelar Budiman³

¹Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

Abstrak

Saat ini, penggunaan audio seperti aplikasi media yang real-time sangatlah dibutuhkan, terutama pada media aplikasi yang menggunakan proses diferensiasi pada data audio, seperti pengkodean yang berbasis konten dan kompresi audio maupun penyetaraan antara speech dan music secara otomatis. Oleh karena itu, diperlukan suatu algoritma yang efisien untuk melakukan segmentasi sinyal audio menjadi speech signal ataupun music signal. Dalam tugas akhir ini, digunakan suatu pendekatan untuk mendeteksi batasan-batasan musik dan mengklasifikasikan speech / music dengan menggunakan suatu algoritma yang dinamakan segmental continuous dynamic programming atau disingkat dengan Segmental CDP.

Algoritma Segmental CDP dapat mengidentifikasi lokasi dari masing-masing bagian musik dan batasan-batasannya berdasarkan berbagai kesamaan segmen dan informasi lokasinya.

Ekstraksi ciri yang digunakan dalam domain waktu diberikan dua pilihan, yaitu : ZCR (Zero Crossing Rate) dan Energi Bit, sedangkan MFCC merupakan ekstraksi ciri dalam domain frekuensi. Pemisahan sinyal campuran berhasil dilakukan dengan menggunakan threshold dari ciri tersebut. Sinyal audio dikategorikan sebagai speech signal jika nilai moving average energy bit \leq nilai maksimum moving average energy bit speech, nilai moving average ZCR \geq nilai minimum moving average ZCR speech, dan nilai moving average MFCC \leq nilai maksimum moving average MFCC speech. Sinyal audio dikategorikan sebagai music signal jika, nilai moving average energy bit \geq nilai minimum moving average energy bit music, nilai moving average ZCR \leq nilai maksimum moving average ZCR music, dan nilai moving average MFCC \geq nilai minimum moving average MFCC music.

Kata Kunci : segmental CDP, speech, music, segmentasi, klasifikasi



Telkom
University

Abstract

Nowadays, the use of audio as the application of real-time media is desperately needed, especially in media applications using the differentiation process on audio data, such as content based encoding and audio compression and equalization between speech and music automatically. Therefore, it is required an efficient algorithm to segment the audio signal into speech signal or music signal. In this final project, we use an approach for detecting music boundaries and classify speech / music by using an algorithm called segmental continuous dynamic programming or shortened by Segmental CDP.

Segmental CDP algorithm can be used to identify the location of each piece of music and their limits based on various similarity segment and location information.

Feature extraction in the time domain is given two options, namely: ZCR (Zero Crossing Rate) and Bit Energy, while the MFCC feature extraction is in the frequency domain. Separation in mixed signals successfully performed using the threshold of the traits. The audio signal is categorized as a speech signal if the value of moving average energy bit \leq maximum value of the moving average bit energy speech, a moving average value of ZCR \geq minimum value of ZCR speech moving average, moving average MFCC and value \leq maximum value of moving average MFCC speech. The audio signal is categorized as a music signal if, the value of moving average energy bit \geq minimum value of moving average energy bit music, moving average ZCR value \leq the maximum value of moving average ZCR music, and moving average MFCC value \leq minimum value of moving average MFCC music.

Keywords : segmental CDP, speech, music, segmentation, classification

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam sebuah rekaman audio, beberapa orang ada yang hanya ingin mendengarkan bagian *speech* saja atau bagian *music* saja. Maka diperlukan sebuah sistem yang bisa mengidentifikasi lokasi dari masing-masing bagian musik dan memberikan batasan-batasan serta mampu mengklasifikasikan apakah suara yang dikeluarkan adalah *speech* atau *music*. Beberapa proses yang mendasar dalam memanipulasi data adalah segmentasi (*segmentation*) dan penggolongan (*classification*) *speech* / *music* yang merupakan langkah pertama dalam memproses data. Preprocessing ini dibutuhkan untuk aplikasi-aplikasi yang membutuhkan pemisahan *speech* yang akurat, seperti transkripsi otomatis pada penyiaran berita, pengenalan suara dan pembicara. Klasifikasi *speech* dan *music* juga berguna untuk aplikasi yang menggunakan proses diferensiasi pada data audio, seperti kompresi dan akualisasi otomatis pada *speech* dan *music*. Hal ini juga dapat digunakan untuk indeks data lainnya, seperti klasifikasi isi video sampai pada pengiringan audio.

Salah satu tantangan pada diskriminasi *speech* / *music* adalah karakteristik yang berbeda antara sinyal *music* itu sendiri. *Speech* terdiri dari pemilihan ciri khas bunyi yang selaras dan sedemikian sehingga dapat direpresentasikan dengan baik ke dalam model yang relatif sederhana. Pada sisi lain, terdapat berbagai macam bunyi pada *music* yang dihasilkan dari berbagai instrumen, sering berasal dari sumber yang bersamaan. Oleh karena itu, pembuatan model untuk representasi akurat dan memuat semua jenis musik sangat rumit. Inilah salah satu pertimbangan solusi algoritma dikembangkan pada proses klasifikasi *speech* / *music*.

Dari hal tersebut, digunakan suatu algoritma Segmental CDP untuk mendeteksi batasan-batasan *speech / music* dan melakukan pengklasifikasian *speech signal* dan *music signal*.

1.2 Perumusan Masalah

Dalam tugas akhir ini akan dibahas beberapa permasalahan, antara lain :

1. Bagaimana sistematika perekaman *live broadcast* pada beberapa siaran radio.
2. Bagaimana proses segmentasi pada rekaman yang sudah dilakukan menjadi segmen-segmen kecil.
3. Bagaimana pengambilan threshold untuk *speech* dan *music*.
4. Bagaimana proses pengklasifikasian hasil segmentasi dengan menggunakan algoritma Segmental CDP.
5. Mengambil kesimpulan dari hasil yang diperoleh.

1.3 Batasan Masalah

Dalam pembahasan tugas akhir ini, masalah dibatasi pada hal-hal berikut :

1. Sinyal masukan yang digunakan merupakan hasil perekaman pada beberapa siaran radio.
2. Format penyimpanan file adalah *.wav
3. Frekuensi *sampling* yang digunakan 44100 Hz.
4. Tidak membedakan jenis musik. (pop, rock, dan lain-lain)
5. Tidak membedakan suara antara laki-laki dan perempuan.
6. Tidak menganalisa suara yang terdapat *background music* misal iklan.

1.4 Tujuan Penelitian

Tujuan penyusunan tugas akhir ini adalah:

1. Memberikan batasan-batasan antara *speech* (suara penyiar radio) dan *music* (lagu yang diputar) pada beberapa rekaman *Radio Broadcast*.
2. Mengetahui pengaruh jumlah segmentasi frame terhadap penentuan nilai *threshold*.
3. Memisahkan antara *speech* (suara penyiar radio) dan *music* (lagu yang diputar) pada beberapa rekaman *Radio Broadcast*.
4. Mengetahui kualitas dari keluaran baik secara subjektif maupun objektif.

1.5 Metodologi Penelitian

Metode penelitian yang digunakan pada tugas akhir ini adalah sebagai berikut:

1. Studi Literatur, berisikan pembahasan teoritis melalui studi literatur dari buku-buku atau jurnal ilmiah yang berkaitan dengan klasifikasi dan segmentasi pada *speech* dan *music*.
2. Perancangan Sistem dan Permodelan untuk melakukan klasifikasi dan segmentasi pada file audio dengan menggunakan algoritma Segmental CDP.
3. Analisa hasil klasifikasi dan segmentasi.
4. Pengambilan kesimpulan dan penyusunan tugas akhir.

1.6 Sistematika Pembahasan

Secara keseluruhan Tugas Akhir ini dibagi menjadi lima bab bahasan, ditambah dengan lampiran dan daftar istilah yang diperlukan. Penjelasan masing-masing bab adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab ini berisi latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi penelitian, dan sistematika penulisan.

BAB II : DASAR TEORI

Berisi tentang teori yang mendukung dan mendasari penulisan tugas akhir ini, yaitu teori tentang klasifikasi dan segmentasi *speech* dan *music*, klasifikasi *music*, algoritma Segmental CDP, dan teori pendukung lainnya.

BAB III : MODEL DAN SISTEM

Bab ini memberikan proses disain simulasi diagram blok untuk segmentasi dan klasifikasi *speech* dan *music*.

BAB IV : ANALISIS HASIL SIMULASI

Bab ini menunjukkan hasil simulasi proses segmentasi dan klasifikasi *speech / music* dengan Matlab versi R2009a dan analisis terhadap simulasi tersebut

BAB V : PENUTUP

Bab ini berisi kesimpulan dan saran dari penelitian yang dilakukan, serta rekomendasi atau saran untuk perbaikan dan pengembangan lebih lanjut.



Telkom
University

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari pengujian dan analisis sistem yang telah dilakukan untuk melakukan proses segmentasi dan klasifikasi *speech / music* menggunakan algoritma segmental CDP, dapat diambil kesimpulan sebagai berikut :

1. Penentuan jumlah segmentasi *frame* akan mempengaruhi nilai parameter. Semakin besar jumlah segmentasi yang digunakan maka akan semakin akurat hasil yang diperoleh.
2. Sinyal audio dikategorikan sebagai *speech* signal jika
 - nilai moving average energy bit \leq nilai maksimum moving average energy bit *speech*
 - nilai moving average ZCR \geq nilai minimum moving average ZCR *speech*
 - nilai moving average MFCC \leq nilai maksimum moving average MFCC *speech*
3. Sinyal audio dikategorikan sebagai *music* signal jika
 - nilai moving average energy bit \geq nilai minimum moving average energy bit *music*
 - nilai moving average ZCR \leq nilai maksimum moving average ZCR *music*
 - nilai moving average MFCC \geq nilai minimum moving average MFCC *music*
4. Pengujian sistem menggunakan MOS mendapat nilai ≈ 4 yang berarti baik.
5. Dari *sample* yang diuji didapatkan nilai MSE sebesar 0,0397 dan akurasi sebesar 87,47 %.

5.2 Saran

Untuk pengembangan tugas akhir ini selanjutnya, dapat dilakukan dengan cara:

1. Menggunakan lebih banyak jenis sampel audio dari berbagai sumber.
2. Menggunakan tools parameter ekstraksi ciri yang lebih lengkap
3. Menggunakan *threshold* secara otomatis sehingga masukan bisa langsung diklasifikasi.
4. Menggunakan sistem perekaman yang lebih bagus agar bisa digunakan secara *real time*.



DAFTAR PUSTAKA

- [1] Itoh, Yoshiaki. Iwabuchi, Akira, and friends. "Automatic Music Boundary Detection Using Short Segmental Acoustic Similarity in a Music Piece". *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2008, Article ID 480786. 2008.
- [2] Lavner, Yizhar. Ruinskiy, Dima. "A Decision-Tree-Based Algorithm for Speech/Music Classification and Segmentation". *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2009, Article ID 239892. 2009.
- [3] Permana, Yunan H.I. "*Usage of Walsh Basis Function on Speech/Non Speech Detection for Increasing it Detection Level of Speech and Non speech*". IT TELKOM. 2010.
- [4] Hamid, Mustofa Abi. "Bunyi atau Suara". <http://mustofaabihamid.blogspot.com>. Juli 2010. [Online] (<http://mustofaabihamid.blogspot.com/2010/07/bunyi-atau-suara.html>)
- [5] Abdi Away, Gunaidi. 2006. *MATLAB Programming*. Informatika Bandung. Bandung.
- [6] Sirait, Renova A.T. "*Musical Sound Separation Based on Binary Time Frequency Masking*". IT TELKOM. 2010.

Telkom
University