

## SEGMENTASI DAN KLASIFIKASI SPEECH DAN MUSIC MENGGUNAKAN ALGORITMA BERBASIS DECISION TREE

Pamriadi P<sup>1</sup>, Bambang Hidayat<sup>2</sup>, Rian Febrian Umbara<sup>3</sup>

<sup>1</sup>Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

---

### Abstrak

Saat ini, penggunaan aplikasi audio banyak dibutuhkan pada media aplikasi yang menggunakan proses diferensiasi pada data audio, seperti pengkodean yang berbasis konten dan kompresi audio maupun penyetaraan antara speech dan musik secara otomatis. Oleh karena itu, diperlukan suatu algoritma yang efisien untuk melakukan segmentasi sinyal audio menjadi speech signal ataupun sinyal musik secara terpisah.

Pada tugas akhir ini, disimulasikan proses segmentasi dan klasifikasi yang dilakukan dengan menggunakan algoritma berbasis Decision Tree. Pada algoritma ini, terdapat dua tahap pengerjaan yaitu tahap pembelajaran dan tahap proses klasifikasi. Masukan yang digunakan dalam analisis adalah sinyal audio yang berasal dari rekaman radio streaming, sedangkan keluaran yang diperoleh berupa potongan speech atau musik.

Pada tahap pembelajaran, diperoleh nilai threshold untuk klasifikasi speech / music menggunakan algoritma berbasis decision tree yaitu untuk speech, threshold short-time energy  $\leq -78,5470$  dB dan mel frequency cepstrum coefficient (MFCC)  $\leq 7,1835$ , untuk musik, threshold short-time energy  $\geq -60,2717$  dB dan Mel Frequency Cepstrum Coefficient (MFCC)  $\geq 7,6848$  serta jumlah segmentasi frame sebanyak 4096 frame dan koefisien moving average sebesar 50.

Pada tahap proses klasifikasi, digunakan 8 sampel audio dengan akurasi minimal 93,97% dan akurasi maksimal 99,99%

Kata Kunci : Decision Tree, Speech, Music, segmentasi, dan klasifikasi

---

### Abstract

Recently, the usage of audio applications are needed in media application that use differentiation of audio data, like content-based coding and audio compression, equalization between speech and music automatically. So an efficient algorithm to segment audio signal into speech signal or music signal separately is needed.

In this final project, segmentation and classification processes using an algorithm based on Decision Tree are applied. In this algorithm, there are two phases of processing, learning phase and clarification phase. The input signal is audio signal from radio streaming recording, while the output signal is part of speech or music.

In the thresholding phase, threshold values for speech/music classification using decision tree-based algorithm are obtained, such that, for speech, threshold Short-time energy  $\leq -78,5470$  dB and mel frequency cepstrum coefficient (MFCC)  $\leq 7,1835$ , for music, threshold Short-Time Energy  $\geq -60,2717$  dB and mel frequency cepstrum coefficient (MFCC)  $\geq 7,6848$  and frame segmentation value is 4096 frame and moving average coefficient is 50.

In the classification process phase, 8 audio samples with minimal accuracy is 93.97% and maximal accuracy is 99,99% are used.

Keywords : decision tree algorithm, speech, music, segmentation, and classification

---

---

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah<sup>[4]</sup>

Beberapa abad lalu, telah diperkenalkan sejumlah data multimedia seperti teks, gambar, *video*, dan *audio*. Manipulasi dan organisasi data yang efisien diperlukan untuk berbagai tugas seperti penggolongan data untuk penyimpanan dan navigasi, proses diferensiasi berdasarkan isi, pencarian informasi yang spesifik, dan lain-lain.

Bagian terbesar dari data adalah *audio* yang banyak digunakan untuk penyiaran, *database*, untuk menjawab pertumbuhan yang cepat mengenai data, maka dikenal sebuah penemuan baru yang dikenal dengan *Audio Content Analysis (ACA)* yang baru-baru ini muncul dengan tujuan meneliti audio dan menyadap informasi isi audio secara langsung dari tiap isyarat akustik dengan tujuan untuk menciptakan “*Table of Contents*”.

Beberapa proses yang dalam manipulasi data adalah segmentasi (*segmentation*) dan penggolongan (*discrimination*) *speech/music* yang merupakan langkah pertama dalam memproses data. praproses 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 *music* dan *speech* 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 diskriminasi *music/speech*.

## 1.2 Tujuan

Tujuan penyusunan tugas akhir ini adalah:

1. Memisahkan antara *speech* (suara / pembicaraan manusia) dan *music* (musik / lagu yang diputar) pada *file* audio dengan format (.wav).
2. Mengetahui pengaruh jumlah segmentasi frame terhadap penentuan nilai *threshold*.
3. Mengklasifikasikan antara *speech* dan musik pada sebuah *file* audio dengan format (.wav).
4. Mengetahui penyebab terjadinya *error* pada tahap klasifikasi.
5. Mengetahui akurasi hasil klasifikasi pada setiap sampel yang diuji.

## 1.3 Rumusan Masalah

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

1. Bagaimana mengubah sebuah file video menjadi sebuah file audio dalam format (.wav).
2. Bagaimana proses segmentasi file audio dengan format (.wav) menjadi sebuah segmen-segmen kecil.
3. Bagaimana pengambilan *threshold* untuk *speech* dan musik.
4. Bagaimana proses pengklasifikasian hasil segmentasi dengan menggunakan metode berbasis *Decision Tree*
5. Mengambil kesimpulan dari hasil yang diperoleh.

## 1.4 Batasan Masalah

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

1. Tidak membedakan jenis music (pop, rock, jazz, dan lain-lain).
2. Tidak membedakan suara antara laki-laki dan perempuan.
3. Tidak menganalisis suara yang terdapat musik latar misal iklan
4. Tidak membahas secara detail tentang konversi *file* video menjadi sebuah *file* audio dengan format (.wav).

---

## 1.5 Metode 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 *Decision Tree*
3. Analisa hasil klasifikasi dan segmentasi.
4. Pengambilan kesimpulan dan penyusunan tugas akhir.

## 1.6 Sistematika Penulisan

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**

Berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penulisan, metode penyelesaian masalah, yang akan digunakan serta sistematika penulisan yang memuat susunan penulisan Tugas Akhir.

### **BAB II Dasar Teori**

Berisi tentang teori yang mendukung dan mendasari penulisan tugas akhir ini, yaitu teori tentang klasifikasi dan segmentasi *speech* dan musik, klasifikasi musik, algoritma *Decision Tree*, dan teori pendukung lainnya.

### **BAB III Perancangan Model Dan Simulasi**

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

### **BAB IV Analisis Hasil Simulasi**

Berisi analisa terhadap hasil yang diperoleh dari tahap perancangan sistem dan simulasi.

---

## BAB V Penutup

Berisi kesimpulan dari analisis yang telah dilakukan, serta rekomendasi atau saran untuk perbaikan dan pengembangan lebih lanjut.



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dari pengujian dan analisis sistem yang telah dilakukan untuk melakukan proses segmentasi dan klasifikasi *speech* / musik menggunakan algoritma berbasis *decision tree*, dapat diambil kesimpulan sebagai berikut :

1. Simulasi segmentasi dan klasifikasi *speech* / musik berdasarkan algoritma *decision tree* berhasil dilakukan dengan menggunakan 8.
2. Pada tahap *Thresholding phase*, penentuan jumlah segmentasi *frame* akan mempengaruhi nilai parameter. Semakin besar jumlah segmentasi yang digunakan maka akan semakin akurat hasil yang diperoleh. Ini dikarenakan frekuensi sampling yang digunakan juga besar yaitu 44100 Hz.
3. Pada tahap *thresholding phase*, penentuan jumlah koefisien *moving average* akan mempengaruhi kerapatan nilai *moving average* parameter. Semakin besar koefisien yang digunakan maka akan semakin rapat pula nilai *moving average* parameter sehingga akan memudahkan dalam pengambilan *threshold*.
4. Pada tahap klasifikasi, tiap sampel akan terjadi *error* karena tidak semua jumlah data *frame* dilakukan proses. Tiap jumlah *frame* sampel akan dibagi dengan jumlah segmentasi *frame* dalam hal ini 4096 *frame* dan dibulatkan ke bawah sehingga sisanya tidak akan dilakukan proses dan dianggap sebagai *error*.
5. *Error* bisa terjadi apabila bagian *speech* masuk dalam bagian segmentasi musik atau sebaliknya. *Error* juga bisa terjadi apabila *speech* masuk dalam

bagian musik atau sebaliknya. Selain itu *error* juga bisa terjadi apabila *speech* / musik tidak termasuk dalam *threshold* yang digunakan dalam hal ini *threshold* yang digunakan yaitu:

- a. Untuk *speech*, *threshold short-time energy*  $\leq -78,5470$  dB dan MFCC  $\leq 7,1835$
  - b. Untuk musik, *threshold short-time energy*  $\geq -60,2717$  dB dan MFCC  $\geq 7,6848$
6. Untuk beberapa percobaan dengan 8 sampel diperoleh akurasi paling kecil yaitu 93,97 % yaitu terjadi pada sampel 8.

## 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 *threshold* secara otomatis sehingga masukan bisa langsung diklasifikasi.
3. Menggunakan sistem perekaman yang lebih bagus agar bisa digunakan secara *real time*.

Telkom  
University

## DAFTAR PUSTAKA

- [1] Anggodo Prihastomo, Antonius Dito. “ *Mengenal FFT-IFFT*”. My Storty Blog. 2010.  
[Online] (<http://deeto88.wordpress.com/2010/09/28/mengenal-fft-iff/> diakses tanggal 21 Desember 2010)
- [2] Basuki, Achmad. Syarif, Iwan. “*Decision Tree*”. Politeknik Elektronika Negeri Surabaya. Surabaya, 2003
- [3] Hamid, Mustofa Abi. “Bunyi atau Suara”. <http://mustofaabihamid.blogspot.com>. Juli 2010.  
[Online] (<http://mustofaabihamid.blogspot.com/2010/07/bunyi-atau-suara.html> diakses tanggal 21 Desember 2010)
- [4] Lavner, Yizhar dan Ruinskiy, Dima. “*A Decision Tree-Based Algorithm for Speech / Music Classification and Segmentation*”. EURASIP Journal on Audio, Speech, and Music Processing. Volume 2009
- [5] Nugroho, Setyo. “*Penerapan Metode Transformasi Fourier Untuk Perbaikan Citra Digital*”. STIKOM Balikpapan, Balikpapan. 2005
- [6] Saeed Khan, Muhammad Kashif. “*Automatic Classification of Speech & Music in Digitized Audio*”. Dhahran, Saudi Arabia. May 2005.
- [7] Yen, Joe. “*Wavelet for Acoustics*”. R98942097.