

# Aplikasi Identifikasi Suara Hewan Menggunakan Metode *Mel-Frequency Cepstral Coefficients* (MFCC)

Ezhar Mega Rison dang <sup>\*1</sup>, Welly Andrian <sup>#2</sup>, Febri Puji Arieska <sup>#3</sup>, Dwi Astuti <sup>#4</sup>, Apri Junaidi <sup>#5</sup>

# Institut Teknologi Telkom Purwokerto  
Jl. D.I. Panjaitan 128, Purwokerto, Indonesia

<sup>1</sup> 15102016@st3telkom.ac.id

<sup>2</sup> 15102038@st3telkom.ac.id

<sup>3</sup> 15102057@st3telkom.ac.id

<sup>4</sup> 15102094@ittelkom-pwt.ac.id

<sup>5</sup> aprijunaidi@ittelkom-pwt.ac.id

Accepted on April 11, 2019

## Abstract

Pengenalan suara berada di bawah bidang komputasi linguistik. Hal ini mencakup identifikasi, pengakuan, dan terjemahan ucapan yang terdeteksi ke dalam teks oleh komputer. Penelitian ini menggunakan *handphone* dan sistem yang dirancang menggunakan suara. Tujuan utama dari penelitian ini adalah menggunakan teknik pengenalan suara untuk mendeteksi, mengidentifikasi dan menerjemahkan suara binatang. Sistem ini terdiri dari dua tahap yaitu pelatihan dan pengujian. Pelatihan melibatkan pengajaran sistem dengan membangun kamus, model akustik untuk setiap kata yang perlu dikenali oleh sistem (analisis *offline*). Tahap pengujian menggunakan model akustik untuk mengenali kata-kata terisolasi menggunakan algoritma klasifikasi. Aplikasi penyimpanan audio untuk mengidentifikasi berbagai suara binatang dapat dilakukan dengan lebih akurat di masa depan.

**Keywords:** Mel-Frequency Cepstral Coefficients (MFCC), Pengenalan Suara, Kuantisasi Vektor

## I. INTRODUCTION

**K**onsep dasar pengenalan suara mengacu pada indentifikasi pembicara, bukan apa yang mereka katakan. Mengenali suara dapat mengurangi tugas menerjemahkan suara dalam sistem yang telah dilatih pada suara orang tertentu untuk melakukan validasi identitas pembicara sebagai bagian dari proses keamanan. Ada daftar sistem pengenalan suara yang tersedia di pasaran. Sistem terbaik dapat dengan mudah mengenali minimal ribuan kata. Namun, sistem ini membutuhkan sesi pelatihan tambahan dimana sistem menjadi disesuaikan dengan vocal dan aksent tertentu. Program ini digunakan oleh banyak orang yang mengalami gangguan fisik. Jadi saat menggunakan komputer mereka dapat bekerja dengan lebih efektif dan efisien. Jenis sistem ini disebut sistem *discrete speech*. Sistem ini membutuhkan pengetahuan tentang pemrosesan sinyal dan pemodelan statistik. Ada dua jenis pemodelan akustik dan pemodelan bahasa. Keduanya merupakan bagian penting dari metode pengenalan suara berbasis statistik saat ini. Model Markov tersembunyi sebagian besar digunakan di banyak daerah. Pemodelan bahasa digunakan secara luas dalam aplikasi pemrosesan bahasa

alami seperti klasifikasi dokumen atau terjemahan mesin statistik.

Tujuan utama yang diikuti untuk pekerjaan ini adalah menggunakan teknik pengenalan ucapan untuk mendeteksi, menerjemahkan dan mengidentifikasi suara binatang dan mengembangkan algoritma untuk pengenalan angka terisolasi seperti yang digambarkan oleh penulis sebelumnya [14]. Kelemahan yang ada di sini adalah bahwa perbandingan akan menjadi tugas yang rumit untuk diikuti. Dalam pengenalan wicara sistem terdiri dari dua tahap pelatihan dan pengujian. Pelatihan melibatkan pengajaran sistem dengan membangun kamus, model akustik untuk setiap kata yang perlu dikenali oleh sistem (Analisis offline). Tahap pengujian melibatkan penggunaan model akustik untuk mengenali kata-kata terisolasi menggunakan algoritma klasifikasi (analisis online). Dengan demikian ada sistem yang diusulkan dalam [12] Dapat mendeteksi setiap suara menggunakan cetak suara uniknya secara akurat, proses ini merekam sinyal suara diproses untuk mendapatkan estimasi spektrum daya. Vektor fitur berasal dari spektrum daya dan plot yang berdekatan serta matriks Töeplitz. Vektor ini telah terbukti menghasilkan cetakan unik yang tidak berulang. tetapi kelemahan yang diangkat di sini adalah Tidak dapat membedakan antara suara pria dan wanita.

Alur kerja pengembangan mengikuti tiga metode yaitu akuisisi suara, analisis bicara dan pengembangan pengguna (UID). Di mana proses akuisisi ucapan dilakukan selama pengembangan kata-kata yang diucapkan oleh seseorang. Di mana pembelajaran tata bahasa dan sintaksis ditunjukkan sebagai bagian dari penguasaan bahasa, penguasaan ucapan terutama berfokus pada pengembangan produksi wicara dan pendekatan wicara pada tahap awal. Proses selanjutnya dalam alur kerja adalah Analitik Bicara yang dapat digunakan untuk memantau dan mengajar agen serta membantu dalam mengelola database yang merupakan perangkat lunak pengenalan suara khusus untuk mengetahui informasi melalui data lisan yang direkam. Tipe ini memiliki perangkat lunak yang mengetahui kata-kata dan menguji ornamen audio untuk mendeteksi stres pada suara pembicara. Menggunakan pendekatan ART [8] penulis mengidentifikasi metode untuk menemukan lagu-lagu burung yang diperkenalkan untuk menemukan keberadaan tunggal saja tetapi hanya dapat menemukan lagu-lagu burung. Metode ini direncanakan menggunakan metode Angular Radial Transform (ART) untuk mengetahui variasi tingkat abu-abu dalam batas gambar di kedua arah sudut dan radial yang selanjutnya terus membentuk metode untuk membentuk gambar spektrogram sehingga untuk mengatasi kekurangan dalam [7] mengembangkan teknik pemrosesan sinyal untuk pengenalan otomatis rempah-rempah burung proses ini sangat membantu karena dengan menggunakan sistem ini kita dapat mendeteksi spesies menggunakan lebih dari satu metode, tetapi hanya lebih dari persentase tertentu yang dapat ditemukan dan untuk variasi lain dari nyanyian burung kebanyakan spesies memerlukan analisis lebih lanjut .

Transformasi radial sudut MPEG-7 yang digunakan dalam [12] disebutkan dalam [4] sebagai alat yang digunakan untuk pengenalan suara, ini digunakan untuk menyediakan header untuk konten audio atau pengindeksan audio yang selanjutnya diklasifikasikan ke dalam dua jenis signifikan berdasarkan teks deskripsi oleh label kategori dan deskriptor kuantitatif menggunakan model probabilistik. Proses ini menyediakan kerangka kerja yang konsisten untuk suara indikasi dan permintaan. Mematuhi berbagai aplikasi yang memiliki komponen audio. Maka hiruk-pikuk adalah salah satu masalah, di mana kualitas suara tidak dapat dicapai sehingga dalam [5] mengekstraksi suara dari hiruk-pikuk yang terletak pada basis spektrum audio (ASB) dan proyeksi spektrum audio (ASP) dari deskriptor audio MPEG-7 adalah metode yang diikuti. Pada mengikuti metode ini ada model positif yang mencapai tingkat akurasi 96% [4],[5]. Ekstraksi fitur, Pengindeksan dan pengambilan dalam database suara binatang ditemukan dalam sistem yang diusulkan di [2] di mana dengan menggunakan metode untuk mengekstraksi algoritma yang mengidentifikasi suara binatang dengan menganalisis kurva seperti lintasan frekuensi waktu dalam sinyal mereka. Metode ini sangat berguna untuk mencari suara binatang untuk berbagai keperluan tetapi kekurangannya di sini adalah kurang bermanfaat atau membantu dalam mencari suara seperti suara binatang dan juga kecepatan pemrosesan permintaan lambat.

Metode pengambilan digunakan dalam audio yang merupakan metode yang diikuti dalam [1],[2] di mana menerapkan analisis audio, pencarian, mesin klasifikasi yang mengklasifikasikan suara berdasarkan kenyaringan, nada, kecerahan dan bandwidth yang meningkatkan proses tetapi kelemahan utama dalam [1] adalah jika dua file audio memiliki fitur akustik yang serupa, mereka sulit untuk dipahami. masalah ini

diselesaikan di [2] di mana sinyal suara dapat direkonstruksi bahkan pada tingkat kesalahan yang rendah sehingga pengambilan audio lebih efektif dan efisien daripada [1] dan metode yang lebih sederhana dapat diharapkan di masa depan [1,2,3]. Pemrosesan audio dan ekstraksi fitur secara rutin merupakan masalah yang berada pada jangkauan luas dan berbagai jenis sinyal audio yang ada. Dalam [10] ekstraksi fitur audio dan skema klasifikasi multi-grup digunakan untuk mengidentifikasi subruang frekuensi waktu yang bias menggunakan teknik local diskriminant base (LDB).

Fitur yang diekstraksi diklasifikasikan sebagai suara buatan dan alami yang diklasifikasikan sebagai suara manusia, instrumental, binatang. Proses ini memiliki tingkat akurasi klasifikasi yang tinggi dan dengan akurasi ini kita harus mendeteksi beberapa suara binatang yang diusulkan dalam [11] untuk mengidentifikasi dan mendeteksi hewan dari vokalisasi mereka yang telah dikembangkan menggunakan ZCR untuk mendeteksi suara dari kebisingan. MFCC digunakan karena kemampuannya untuk mewakili spektrum bicara dalam bentuk yang ringkas. DWT digunakan untuk klasifikasi pola suara [10],[11]. Sistem pengenalan speaker waktu nyata diusulkan yang dapat mendeteksi satu orang dengan akurasi sangat tinggi [13] ini dilakukan dengan mengubah sinyal audio akustik menjadi bentuk yang dapat dibaca komputer menggunakan ekstraksi fitur dan pencocokan fitur.

Pendekatan yang dilakukan di sini adalah mengenali suara yang dibuat oleh hewan yang berbeda dan mencari tahu menggunakan algoritma yang menyaring suara dan menemukan suara binatang yang terisolasi dan menampilkan nama hewan tersebut. Model akustik juga digunakan yang berisi representasi statistik dari setiap kata yang disebut fonem Sistem kami dapat digunakan dalam proses survei hewan, aplikasi buku audio penyimpanan suara untuk mengidentifikasi suara-suara binatang yang berbeda di masa depan dengan lebih akurat.

## II. LITERATURE REVIEW

Dalam [1] penulis menjelaskan, Kecepatan dan pertumbuhan kapasitas di komputer dan jaringan memungkinkan dimasukkannya audio sebagai tipe data dalam aplikasi komputer saat ini. Klasifikasi dan pencarian audio berdasarkan fitur perseptualnya, bukannya koleksi byte yang buram seperti bidang primitif yaitu nama, ukuran, format file. Pendekatan utama akan menerapkan analisis Audio, pencarian, mesin klasifikasi yang mengklasifikasikan suara berdasarkan kenyaringan, nada, kecerahan, dan bandwidth. Dengan melakukan ini, audio tertentu dapat dengan mudah diidentifikasi dan diambil. Tetapi jika dua file audio memiliki fitur akustik yang serupa, mereka sulit untuk dibedakan. Di sini ia mengakses file audio yang ada yang hanya ada di database dan bukan yang diwakili dalam database. Setiap suara memiliki suara latar belakang yang harus dipisahkan dari suara utama, untuk psikologi Gestalt ini atau pemrosesan sinyal non-persepsi digunakan. Karenanya ini meningkatkan proses klasifikasi, pencarian dan pengambilan audio daripada metode yang digunakan sebelumnya.

Dalam [2] Rolf Bardeli mengusulkan sistem untuk pencarian kesamaan dalam database suara binatang dengan menggunakan algoritma yang ada. Dengan menggunakan algoritma ekstraksi fitur yang mengidentifikasi suara binatang dengan menganalisis lintasan frekuensi waktu seperti kurva dalam sinyal mereka. Dengan mengikuti ini, suara basis data hewan akan sangat membantu untuk menemukannya. Tetapi kurang membantu dalam mencari suara binatang seperti suara, kecepatan pemrosesan permintaan lambat. Dalam sistem yang lebih lama, Hidden Markov Model (HMM) digunakan untuk pengakuan. Dalam hal ini, penulis menggunakan algoritma ekstraksi fitur atau mendapatkan jari yang ringkas seperti resep suara yang akan dicari atau diidentifikasi. Sinyal audio yang dihasilkan oleh hewan menciptakan kurva seperti spektrum yang dapat digunakan untuk mengidentifikasi hewan tertentu dengan menggunakan teknik pemrosesan gambar. Tujuan utamanya adalah untuk secara unik mengenali dan menemukan suara yang serupa di dalam basis data.

Untuk menerapkan metode ini, pengindeksan harus dilakukan yang merupakan proses yang memakan waktu sehingga pemrosesan pada basis data besar membutuhkan waktu. Jadi ekstraksi fitur, pengindeksan dan pengambilan dalam database suara binatang dapat dilakukan dengan menggunakan metode ini. Guodong Guo dan Stan Z. Li pada awalnya menceritakan tentang klasifikasi yang efektif dan pengambilan audio. Pendekatan utama di sini adalah Fitur perseptual, fitur Mel-cepstral dan kombinasinya. Koefisien cepstral menangkap bentuk spektrum frekuensi audio. Penulis menggunakan mesin dukungan vektor (SVM) dengan pengenalan pohon biner yang digunakan untuk mengatasi masalah klasifikasi audio yang merupakan

algoritma yang baru diusulkan untuk pengenalan pola. Permintaan data audio yang meningkat pesat untuk metode modern yang memungkinkan klasifikasi berbasis konten yang efisien dan terkomputerisasi serta pemulihan audio dari basis data. SVM biasanya digunakan untuk mengklasifikasikan audio yang menjadi milik satu kelas, Tapi di sini fungsi dari SVM diperluas untuk memecahkan masalah klasifikasi multi-kelas yang selama tinjauan eksperimental memiliki tingkat kesalahan sangat kurang dibandingkan dengan metode tradisional Suatu karya terbaru yang menakjubkan dilakukan oleh Wold et al. yang disebut sebagai Muscle Fish digunakan yang membedakannya dari pekerjaan pengambilan audio sebelumnya dalam kemampuannya.

Dalam sistem Muscle Fish, berbagai fitur, seperti kenyaringan, kecerahan, nada, timbre digunakan untuk mewakili suara. Euclidean yang dinormalisasi juga dikenal sebagai Mahalanobis yang merupakan jarak dan aturan tetangga terdekatnya (NN) yang digunakan untuk mengklasifikasikan suara kueri ke dalam kelas suara dari database-nya. Dalam [4] Michael Casey meneliti bahwa pengindeksan efisien file audio dalam basis data Penggunaan kerangka kerja yang kuat yang diterima secara universal. Pendekatan utama dalam makalah ini adalah Alat pengenalan suara MPEG-7 menyediakan header untuk pengindeksan audio. Deskripsi dibagi menjadi dua jenis: deskripsi berbasis teks oleh label kategori dan deskripsi kuantitatif menggunakan model probabilistik. Alat pengenalan suara MPEG-7 menyediakan antarmuka terpadu untuk pengindeksan otomatis audio menggunakan kelas suara terampil dalam kerangka kerja pengenalan prototipe. Dengan melakukan ini, ia menyediakan kerangka kerja yang konsisten untuk pengindeksan dan permintaan suara yang dapat disesuaikan untuk berbagai aplikasi yang memiliki komponen audio. Dengan demikian Sound Recognition Classifier berfungsi sebagai header yang menyediakan deskripsi tentang audio. Ini dapat digunakan untuk menemukan audio.

Agar metodologi ini berfungsi, kita harus menyimpan fungsi dasar dan parameter HMM, model referensi dalam database. Menggunakan fitur-fitur ini header dibuat untuk pelabelan dan ekstraksi audio yang efisien menggunakan MPEG-7. Dalam metode [5] Hyoung-Gook Kim, Nicolas Moreau, dan Thomas Sikora memperkenalkan teknik untuk mengevaluasi efisiensi sistem pengindeksan dan pengambilan audio. Ekstraksi suara dari kebisingan. Teknik ini menggunakan skema deskripsi level rendah dan level tinggi. Untuk sinyal audio spektrum tingkat rendah dihasilkan dan digunakan. Untuk pemodelan fitur dimensi rendah tingkat tinggi digunakan. Pendekatan utama adalah memperkirakan efektivitas metode pengindeksan dan pengambilan audio berdasarkan basis spektrum audio (ASB) dan proyeksi spektrum audio (ASP) dari deskriptor audio MPEG-7. Kelebihan utama adalah suara Audio yang mengklasifikasikan ke dalam kelas suara yang dipilih secara real time dengan akurasi 96%. Namun Akurasi berkurang dengan kesamaan suara. Dalam naskah ini, mereka menerapkan deskriptor audio ASB dan ASP MPEG-7 untuk dua metode pengenalan, pengenalan speaker dan kategorisasi suara dan sistem pengambilan. Pengenal speaker, diuji dengan speaker wanita dan pria, menghasilkan tingkat identifikasi yang tinggi. Untuk evaluasi, MFCC biasa dengan fitur delta dan delta ganda diekstraksi.

Hasil uji coba menunjukkan bahwa tingkat pengenalan menggunakan fitur ASP dimensi lebih kecil dari vektor karakteristik MFCC dimensi. tipe ASP menggunakan fungsi berbasis ICA memverifikasi pembicara dan tindakan pengenalan gender yang lebih baik daripada fitur NASE dan jenis proyeksi berbasis PCA. Dengan demikian kita dapat mengidentifikasi ekstraksi suara di bawah berbagai lingkungan dan cukup umum untuk menggambarkan berbagai kelas suara. Michael Clausen dan Frank Kurth mengusulkan suatu sistem dalam [6] pada penerapan pendekatan untuk pengenalan musik berbasis indeks cepat yang banyak digunakan dalam pengambilan informasi musik (MIR). Gagasan utama di balik ini adalah untuk mencari dan mengidentifikasi file musik dalam database dengan akurasi tinggi. Pendekatan yang dicoba oleh penulis ini adalah data skor polifonik dan identifikasi bahan audio modulasi kode pulsa dari bentuk gelombang akustik yang diberikan. Kelebihan utama di sini adalah Ukuran indeks pencarian kami sangat kecil. Sama, pengindeksan dan kinerja pengambilan sangat tinggi. Tetapi ini kompleks. Penulis menggunakan fitur notify dan audentify untuk tujuan ini. Oleh karena itu Pencarian dan pengambilan didasarkan pada musik berbasis skor dan identifikasi audio PCM dilakukan di sini.

Dalam [7] Panu Somervuo et al. mengembangkan Pengembangan teknik pemrosesan sinyal untuk pengenalan otomatis spesies burung. Pendekatan utama yang dilakukan oleh penulis adalah pemodelan sinusoidal untuk burung-burung dengan nada tinggi. Parameter Mel-cepstrum mengenali spesies burung langsung dari suku kata (elemen) yang merupakan blok bangunan dari nyanyian burung. Di mana Lebih dari satu metode hadir untuk mendeteksi nyanyian burung dari spesies yang berbeda tetapi persentase pengakuan yang benar lebih dari 70%, tetapi rata-rata hanya sekitar 50% tergantung pada metode yang diikuti. Dengan

demikian penulis menyimpulkan bahwa spesies burung tertentu dapat diprediksi menggunakan metode yang tersedia saat ini, sedangkan variasi lagu burung dari sebagian besar spesies memerlukan analisis lebih lanjut.

Deepika M, Nagalinga Rajan menemukan cara [8] untuk menciptakan sistem untuk mengidentifikasi lagu-lagu burung. Pendekatan utama oleh penulis adalah deskriptor MPEG-7 Angular Radial Transform (ART) yang digunakan untuk gambar digunakan untuk menjelaskan variasi tingkat abu-abu dalam wilayah gambar dalam urutan angular dan radial, untuk mengekstraksi fitur bentuk dari gambar spektrogram. Kelebihannya di sini adalah Dapat digunakan untuk mengidentifikasi spesies burung yang berbeda tetapi kelemahan utamanya adalah hanya lagu-lagu burung yang diidentifikasi. Penulis akhirnya menyimpulkan bahwa deskriptor baru diusulkan untuk mengidentifikasi spesies burung dalam burung yang direkam.

Yoshio Ikeda, Yohei Ishii pada tahun 2008 mengusulkan model [9] untuk menemukan perubahan suara sapi di bawah dua kondisi 1. Saat lapar 2. Saat dipisahkan dari betis. Dimana pendekatan utama adalah untuk menemukan varians dihitung dan bagian varians tertinggi secara spektral ditemukan oleh model linear predictive coding (LPC). Kelebihan yang ditemukan dalam makalah ini adalah informasi mengenai frekuensi resonansi berbeda yang ada dalam suara ditemukan tetapi ini tidak akurat yang terletak sebagai kelemahan utama dalam makalah ini. Akhirnya penulis menyimpulkan bahwa Ditemukan bahwa kerusakan psikosomatis dari sapi di bawah divisi dari betisnya menghasilkan suara dengan frekuensi resonansi yang lebih rendah daripada suara-suara dalam kondisi kelaparan. Karthikeyan Umapathy et al menemukan [10] solusi untuk pemrosesan Audio dan ekstraksi fitur pada rentang yang lebih luas dan kategori sinyal audio.

Cara penulis mendekati masalah ini adalah dengan mengekstraksi fitur audio dan skema klasifikasi multi-grup digunakan untuk mengidentifikasi subruang frekuensi-waktu diskriminatif menggunakan teknik local diskriminant base (LDB). Fitur yang diekstraksi diklasifikasikan sebagai suara buatan dan alami yang diklasifikasikan sebagai suara hewan, instrumental manusia. Kelebihan dalam makalah ini adalah Tingkat akurasi klasifikasi yang tinggi ditemukan. Penulis akhirnya menyimpulkan bahwa skema klasifikasi audio berbasis LDB yang mencakup berbagai sinyal audio disajikan. Akurasi klasifikasi tinggi dicapai dengan menggunakan metodologi yang diusulkan.

Dalam [11] penulis menjelaskan bagaimana mereka mengidentifikasi suara yang berbeda dari berbagai frekuensi vokalisasi. Di mana ZCR digunakan untuk mendeteksi suara dari kebisingan. MFCC digunakan karena kemampuannya untuk mewakili spektrum bicara dalam bentuk yang ringkas. DWT digunakan untuk klasifikasi pola suara. Kelebihan yang ditemukan di sini adalah dapat mendeteksi beberapa suara binatang dan kelemahan terbesar di sini adalah kualitas sinyal minimal dan Memerlukan sinyal berkualitas tinggi untuk pengakuan dan akhirnya penulis menyimpulkan bahwa Identifikasi dan Deteksi hewan dari vokalisasi mereka dapat ditemukan melalui metode ini. Masalah yang dibahas dalam [12] adalah untuk mengidentifikasi setiap suara yang unik dengan cetak suara yang unik. Dan pendekatan yang diikuti oleh penulis adalah merekam sinyal suara yang diproses untuk mendapatkan estimasi spektrum daya. Vektor karakteristik adalah tiruan dari spektrum daya dan plot di dekatnya dan matriks Töeplitz. Vektor ini telah terbukti menghasilkan cetakan unik yang tidak berulang. Kelebihannya di sini adalah dapat mendeteksi setiap suara menggunakan cetak suara uniknya secara akurat tetapi tidak dapat membedakan antara suara pria dan wanita dan dapat dengan mudah dilewati. Tinjauan singkat yang diberikan dalam makalah ini memmanifesasikan pendekatan baru untuk voiceprint sebagai sarana untuk deskripsi data yang sederhana.

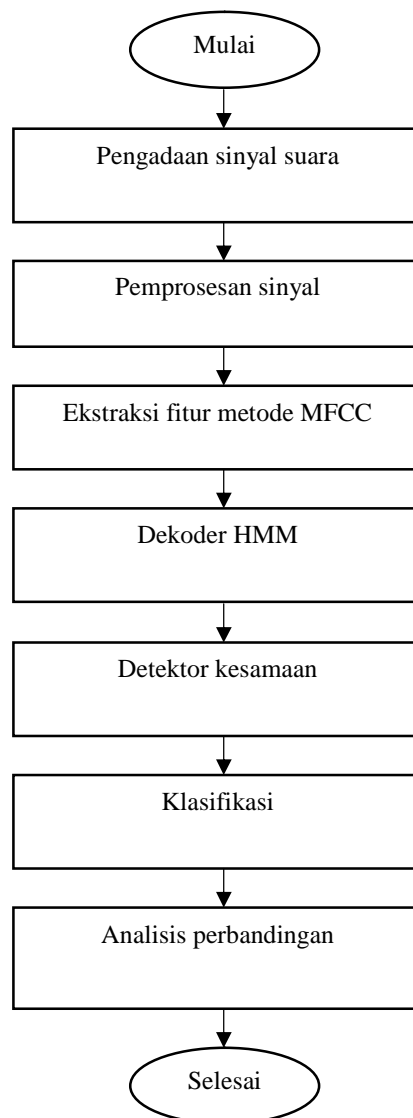
Roma Bharti mengusulkan sistem [13] yang merupakan pembicara waktu nyata bahwa sistem pengenalan yang telah dilatih untuk pembicara tertentu dan memverifikasi pembicara. Pendekatan yang diikuti di sini adalah dengan mengubah sinyal audio akustik menjadi bentuk yang dapat dibaca komputer. Menggunakan ekstraksi fitur dan pencocokan fitur. Kelebihan yang ditemukan dalam makalah ini adalah dapat dengan mudah mendeteksi satu orang dengan akurasi yang sangat tinggi dan kekurangannya di sini adalah membutuhkan sinyal suara yang baik. Dapat mendeteksi hanya satu orang suara pada suatu waktu. Penulis akhirnya menyimpulkan bahwa sistem ini mengenali speaker waktu nyata dengan bantuan database yang tersimpan dan dapat diperluas dengan jumlah pengguna yang lebih banyak.

Pada [14] Yu-Hsiang Bosco Chiu dkk meneliti dan mengangkat masalah yang lebih baik dalam akurasi pengenalan di Kehadiran kebisingan latar belakang daripada ekstraksi fitur MFCC yang diakui dan pendekatan yang diikuti di sini adalah kriteria MMI (Mutual Information Information) maksimum yang digunakan untuk menentukan sinyal dari kebisingan. Kelebihan yang ditemukan dalam makalah ini adalah dapat mengenali suara bahkan selama adanya kebisingan dan kelemahan yang ditemukan adalah

Menggunakan database yang kompleks untuk perbandingan dan akhirnya Memiliki akurasi yang lebih tinggi daripada metode pengenalan suara tradisional.

### III. RESEARCH METHOD

Tahapan penelitian berguna untuk mendapatkan informasi yang dapat menjawab pertanyaan penelitian. Adapun tahapan penelitian seperti berikut.



Gambar 3.1 Tahapan Penelitian

Tahapan – tahapan pada Gambar 3.1 akan dijelaskan secara rinci pada penulisan dibawah ini:

**a. Pengadaan sinyal suara**

Sinyal diperoleh dengan menggunakan mikrofon bebas distorsi, kemudian dari sinyal yang dihasilkan informasi yang tidak diinginkan dihapus untuk menjaga sifat suara unik dari suara hewani yang harus diidentifikasi selama pemrosesan sinyal.

**b. Pemrosesan sinyal**



Pada tahapan ini sinyal suara yang sudah diambil kemudian diproses dan direkan pada frekuensi 44100Hz dan disimpan dalam format 8 bit dan dinormalisasi sehingga amplitude sinyal berada dalam kisaran -1 hingga 1.

Penggunaan filter noise, kisi-kisi noise spektral, dan analisis sinyal digunakan untuk hapus noise selama pemrosesan sinyal dan dapatkan parameter anti noise untuk mendapatkan kualitas sinyal yang baik dari latar belakang bisung yang ada di sinyal audio yang direkam.

### 1. Spektrum atau Kontinum daya

Daya Kontinum atau spektrum daya dari suatu rangkaian waktu menggambarkan alokasi komponen frekuensi yang membentuk sinyal itu. Menurut analisis Fourier, setiap sinyal penting dapat didekomposisi menjadi sejumlah Frekuensi atau variasi frekuensi pada rentang yang terus menerus. Rata-rata aritmatika dari sinyal tertentu termasuk kebisingan yang dianalisis dalam ketentuan konten frekuensinya disebut spektrumnya. Ketika energi sinyal diintensifkan di sekitar interval waktu yang terbatas, khususnya jika energi totalnya terbatas, kita dapat menghitung kerapatan spektral energi. Sinyal yang direkam diproses untuk mendapatkan spektrum.

### 2. Spektrogram

Spektrogram adalah ilustrasi dari rangkaian frekuensi dalam suatu bunyi karena mereka berbeda dengan instance atau variabel lainnya. *Spectrograms* kadang-kadang disebut *voiceprints spektral*. *Centroid spektral* sinyal adalah rata-rata fungsi kerapatan spektralnya, yaitu kejadian yang membagi alokasi menjadi dua bagian yang sama. Frekuensi batas spektral sinyal adalah tambahan dari konsep sebelumnya untuk setiap kuantitas di tempat dua bagian yang sama. Kepadatan spektral adalah fungsi frekuensi, bukan peran waktu. Namun, kepadatan spektral dari jendela berukuran kecil dari sinyal yang lebih panjang dapat dihitung, dan diplot terhadap waktu yang dikaitkan dengan jendela. Grafik semacam itu disebut *spektrogram*. Ini adalah sumber dari sejumlah teknik studi spektral seperti transformasi Fourier waktu-pendek dan wavelet. Spektrogram dapat digunakan untuk memeriksa hasil melewati sinyal analisis melalui pemeriksa sinyal seperti filter untuk memastikan kinerjanya. Spektrogram biasanya dibentuk dalam dua cara yang diperkirakan sebagai bank saringan yang menghasilkan urutan band pass atau direncanakan sebelumnya dari sinyal instan menggunakan FFT.

### c. Ekstraksi Fitur metode MFCC

paling umum digunakan dalam sistem pengenalan suara otomatis untuk mengekstraksi fitur adalah koefisien cepstral *Mel-Frequency* yang digunakan untuk mendeteksi komponen sinyal audio yang berguna untuk secara akurat mendeteksi konten linguistik dan menghilangkan lingkungan suara, distorsi, dll. Suara yang dihasilkan oleh manusia, hewan dan spesies lain disaring oleh saluran vokal yang memberikan bentuk tertentu pada suara. Bentuk ini menentukan mana yang diproduksi. Jika kita dapat menentukan bentuk ini dengan tepat, kita bisa mendapatkan gambar yang tepat dari fonem yang sedang dibuat. Bentuk saluran vokal itu sendiri membentuk paket dari spektrum daya waktu singkat, dan tujuan MFCC adalah untuk secara tepat mencirikan paket ini. Sinyal audio terus berubah, sehingga untuk memudahkan kita anggap bahwa pada interval waktu kecil sinyal audio tidak banyak berubah. Ini kita membagi sinyal menjadi frame (30-50 ms frame per sinyal). Jika frame jauh lebih kecil maka kami tidak dapat memperoleh sampel yang cukup untuk mendapatkan estimasi spektral yang konsisten, jika tidak, sinyal akan berubah terlalu banyak dalam frame. Langkah selanjutnya adalah memperkirakan spektrum daya setiap frame. Untuk ini kami menggunakan bank Mel filter. Filter pertama sangat tipis dan memberikan petunjuk tentang bagaimana sebagian besar energi ada mendekati nol Hertz. Ketika frekuensi semakin tinggi, filter semakin tebal seiring bertambahnya variasi kita. Kami hanya memperhatikan sekitar berapa banyak energi yang terjadi di setiap tempat. Level Mel memberitahu kita secara tepat bagaimana menempatkan bank filter dan seberapa lebar menempatkannya. Langkah terakhir adalah menghitung DCT dari energi bank filter log. Ada dua alasan mengapa ini dilakukan. Pertama, bank filter semuanya saling tumpang tindih; energi bank filter sangat saling terkait satu sama lain. DCT de mengaitkan energi-energi ini yang matriks kovarians diagonalnya dapat digunakan untuk merepresentasikan fitur-fitur dalam missal . Klasifikasi HMM. Koefisien DCT yang tinggi melambangkan perubahan cepat dalam energi bank filter dan ternyata perubahan ini menurunkan kinerja ASR.

### d. Dekoder

HMM HMM (Hidden Markov Model) diimplementasikan dalam pengenalan ucapan karena sinyal ucapan dapat direpresentasikan sebagai sinyal diam secara bertahap. Sebuah pidato contoh singkat dapat dilihat sebagai proses stasioner. Komunikasi verbal dapat dipertimbangkan sebagai model Markov untuk banyak tujuan stokastik. Dasar lain mengapa HMM disukai adalah karena dapat dilatih secara otomatis, tidak

rumit dan realistis secara komputasi untuk digunakan. Dalam pengenalan ucapan, model Markov yang tersembunyi akan menghasilkan serangkaian vektor bernilai nyata  $n$ -dimensional. Vektor berisi koefisien cepstral, yang diperoleh dengan mengambil transformasi Fourier dari jendela waktu kecil dan de menghubungkan spektrum menggunakan transformasi kosinus, kemudian mempertimbangkan koefisien pertama. Model Markov tersembunyi akan cenderung memiliki di setiap negara bagian distribusi statistik yang merupakan kombinasi dari kovarian *Gaussians diagonal*, yang akan memberikan kemungkinan untuk setiap vektor eksperimental. Setiap ujaran, atau fonem, akan memiliki alokasi keluaran yang berbeda. Model Markov tersembunyi untuk serangkaian ujaran atau fonem disusun dengan menggabungkan model Markov tersembunyi yang diajarkan secara individual untuk ujaran dan fonem yang terpisah.

#### e. Klasifikasi

Here classifier digunakan untuk mengklasifikasikan suara binatang yang berbeda berdasarkan karakteristik akustik dan nilai cepstral mereka. Di sini kuantisasi vektor digunakan untuk memplot nilai dan mengklasifikasikan nilai berdasarkan posisi vektor pada grafik. Pada penelitian ini menggunakan kuantisasi dalam pemrosesan sinyal.

Kuantisasi vektor adalah metode kuantisasi tradisional dalam pemrosesan sinyal yang digunakan untuk pemodelan fungsi kerapatan peluang dengan alokasi vektor mode. Kemampuan pencocokan kepadatan vektor kuantisasi sangat bagus, terutama untuk mengidentifikasi kepadatan informasi besar dan dimensi tinggi. Karena titik-titik informasi diwakili oleh indeks centroid terdekat mereka, informasi yang sering muncul memiliki kesalahan yang rendah dan juga data yang tidak akurat yang langka. Di sini untuk mengenali suatu pola, sebuah codebook dibuat untuk masing-masing menggunakan vektor akustik dari pengguna.

#### f. Analisis

Semua metode yang disebutkan di atas digunakan dalam penelitian ini seperti MFCC, DTW, Power spectrum, HMM, Vector quantization digunakan untuk membandingkan dan mengidentifikasi suara binatang secara real time. Sistem yang disebutkan di atas adalah kombinasi dari beberapa metodologi untuk mendapatkan akurasi yang lebih tinggi dalam mengenali suara dalam situasi kehidupan

### IV. RESULTS AND DISCUSSION

Semua metode yang disebutkan seperti MFCC, DTW, Power spectrum, HMM, Vector kuantization digunakan untuk membandingkan dan mengidentifikasi suara binatang secara real time. Sistem yang disebutkan di atas adalah kombinasi dari beberapa metodologi untuk mendapatkan akurasi yang lebih tinggi dalam mengenali suara dalam situasi kehidupan nyata

### V. Conclusion

Isolated Word Recognition System for Animal voice recognition identification System diajukan dalam penelitian ini, dengan tujuan untuk dapat mengenali suara binatang di lingkungan sekitar. Dengan sistem yang diusulkan ini, kita dapat mengidentifikasi hewan dengan suaranya dengan membandingkannya dengan sampel yang sudah ada dalam database. Sistem ini dapat digunakan untuk survei hewan, ilmu kedokteran hewan, dan penelitian zoologi. Di masa depan, kami bermaksud menggunakan dasar dari sistem ini untuk mendeteksi perubahan suara hewan dalam kondisi psikologis mereka yang berbeda untuk memahami keadaan psikologis hewan dan untuk digunakan dalam pelatihan pengklasifikasi, dan karenanya kami mengantisipasi untuk dapat mencapai tingkat akurasi yang lebih tinggi dalam mengenali suara berbagai hewan yang ada di dunia.

### ACKNOWLEDGMENT

Terimakasih kepada Lembaga Penelitian dan Pengabdian Pada Masyarakat (LPPM) Institut Teknologi Telkom Purwokerto. Terimakasih kepada Bapak Apri Junaidi selaku dosen pengenalan suara yang telah memotivasi.



## REFERENCES

- [1] Erling Wold, Thom Blum, and Douglas Keislar, and James Wheaton, "Content-Based Classification, Search and Retrieval of Audio" IEEE Multimedia, vol. 3, no. 3, pp.27 – 36, 1996.
- [2] Rolf Bardeli, "Similarity Search in Animal Sound Databases", IEEE Multimedia, vol. 11, no.1, pp. 68-76, Jan 2009.
- [3] Guodong Guo and Stan Z. Li, "Content-Based Audio Classification and Retrieval by Support Vector Machines", IEEE Neural Networks, vol. 14, no. 1, Jan 2003.
- [4] Michael Casey, "MPEG-7 Sound-Recognition Tools", IEEE Trans. Circuits and systems for video technology, vol.11, no.6, June 2001.
- [5] Hyoung-Gook Kim, Nicolas Moreau, and Thomas Sikora, "Audio Classification Based on MPEG-7 Spectral Basis Representations", IEEE Tans. Circuits and systems for video technology, vol. 14, no. 5, May 2004.
- [6] Michael Clausen and Frank Kurth, "A Unified Approach to Content-Based and Fault-Tolerant Music Recognition", IEEE Trans. Multimedia, vol. 6, no. 5, Oct 2004.
- [7] Panu Somervuo, Aki Härmä, and Seppo Fagerlund, "Parametric Representations of Bird Sound for Automatic Species Recognition", IEEE Trans. Audio, Speech and Language processing, vol. 14, no. 6, Nov 2006.
- [8] Deepika M and Nagalinga Rajan, "Automatic Identification of Bird Species from the Recorded Bird Song Using ART Approach", Int. Conf. on innovations in engg., vol. 3, no. 3, Mar 2014.
- [9] Yoshio Ikeda and Yohei Ishii, "Recognition of two psychological conditions of a single cow by her voice", Journal in Computers and Electronics in Agriculture , vol. 62. no. 1, pp. 67-72 • June 2008
- [10] Karthikeyan Umapathy, Sridhar Krishnan and Raveendra K. Rao, "Audio Signal Feature Extraction and Classification Using Local Discriminant Bases", IEEE Trans. Audio, Speech and Language processing, vol. 15, no. 4, May 2007.
- [11] A. D. Mane, Rashmi R. A, and S. L. Tade, "Identification & Detection System for Animals from their Vocalization", Int. Jour. Advanced computer research, vol. 3, no. 3, Sept 2013.
- [12] Khalid Saeed, "Sound and Voice Verification and Identification A Brief Review of Töeplitz Approach", 7th Conference Znalosti, pp. 22-27, 2008.
- [13] Roma Bharti and Priyanka Bansal, "Real Time Speaker Recognition System using MFCC and Vector Quantization Technique", Int. Jour. Computer Applications, vol. 117, no. 1, May 2015.
- [14] Yu-Hsiang Bosco Chiu, Bhiksha Raj and Richard M. Stern, "Learning-Based Auditory Encoding for Robust Speech Recognition", IEEE Trans. Audio, Speech and Language processing, vol. 0, no. 0, 2011.