

1. Pendahuluan

Pada bagian ini berisi empat sub-bagian, yaitu Latar Belakang, Topik dan Batasannya, Tujuan, dan Organisasi Tulisan. Berikut adalah mengenai penjelasannya.

Latar Belakang

Pengenalan pembicara adalah salah satu teknologi biometrik untuk mengenali siapa yang sedang berbicara. Suara memiliki keunikan masing-masing, seperti pada sidik jari, DNA, dan iris mata, dikarenakan karakteristik suara tiap manusia berbeda-beda karena dipengaruhi faktor karakter fisik kerongkongan [1], sehingga ketepatan pada saat pengklasifikasian sebaiknya tinggi [2]. Terdapat dua kategori di dalam Pengenalan pembicara, yaitu *Text-Independent* yang berarti system akan mengabaikan kata dan bahasa yang diucapkan dan *Text-Dependent* yang akan memroses juga kata yang ada untuk dicocokkan [3].

Pengenalan pembicara berbeda dengan model autentifikasi yang lainnya dikarenakan pada metode ini contoh suara yang didapat dengan cara dinamis, misalnya direkam untuk beberapa detik [1]. Pengenalan pembicara bisa menjadi pembeda atau identitas dikarenakan karakter suara tiap orang berbeda-beda. Karena karakter suara dipengaruhi oleh bentuk fisik dan logat masing masing orang [1].

Pada proses Pengenalan Pembicara terdapat beberapa tahap yang dilakukan sehingga akhirnya suara itu bisa dikenali. Tahap pertama ada ekstraksi fitur. Seiring berjalannya waktu, metode untuk melakukan ekstraksi fitur selalu berkembang sehingga menjadi lebih baik. Metode ekstraksi fitur yang telah ditemukan saat ini ada Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Cepstral Coefficients (LPCC), RASTA Filtering, dan Probabilistic Linear Discriminate Analisis (PLDA) [4]. Metode yang akan digunakan pada tahap ekstraksi fitur adalah MFCC karena telah terbukti memiliki performa yang terbaik dibanding fitur ekstraksi yang lain [4]. MFCC memroses sinyal suara seperti pendengaran manusia, karena dia memproses suara itu hampir seperti telinga manusia. Telinga manusia memiliki filter yang berbeda-beda, ada filter saat nada tinggi atau nada rendah. Lalu pada MFCC, terdapat filter yang berbeda beda juga dan memprosesnya secara logaritmik.

Sebelum tahap ekstraksi, akan dilakukan metode Short Time Zero Crossing Rate (STZCR) yang berfungsi untuk mendeteksi bagian pengucapan konsonan/*unvoiced* yang dianggap sebagai *noise* dikarenakan bentuk gelombangnya yang non-periodik dan suaranya acak yang disebabkan pada saluran kerongkongan [5] sehingga diharapkan bisa meningkatkan akurasi dibanding tidak dilakukan pemotongan.

Setelah fitur diekstraksi dari suara, selanjutnya akan diproses pada tahap pemodelan & klasifikasi. Metode yang biasa digunakan pada Speaker Recognition adalah Gaussian Mixture Model (GMM), Dynamic Time Warping (DTW), dan Support Vector Machine (SVM). Pada pengujian yang dilakukan [6], GMM adalah model yang memiliki akurasi tertinggi, sehingga pada penelitian ini menggunakan metode GMM.

Topik dan Batasannya

Sistem ini dibuat dengan menggunakan metode Short Time Zero-crossing Rate untuk mengurangi bagian *unvoiced* atau konsonan pada suara manusia yang ada pada 71 kelas dataset *VoxForge* sebelum melalui tahap ekstraksi fitur Mel Frequency Cepstral Coefficients dan klasifikasi menggunakan Gaussian Mixture Models lalu dibandingkan dengan metode yang tidak dilakukan pemotongan.

Tujuan

Dengan dibuatnya penelitian ini, dapat mengetahui bagaimana performansi yang dihasilkan oleh pengurangan bagian *unvoiced* pada dataset suara berdasarkan Short Time Zero Crossing Rate sebelum dilakukan ekstraksi fitur Mel Frequency Cepstral Coefficient sebagai pengenalan pembicara pada klasifikasi menggunakan Gaussian Mixture Models dibanding dengan yang tidak dilakukan pemotongan.

Organisasi Tulisan

Laporan ini dibagi menjadi beberapa bagian untuk menjelaskan hasil dari pengujian. Pada bagian 1 dijelaskan latar belakang, batasan dan tujuan penelitian ini dibuat. Pada bagian 2 dijelaskan metode yang digunakan dan bagian 3 menjelaskan penerapannya pada sistem. Bagian 4 adalah hasil pengujian yang dilakukan dan kesimpulan dituliskan pada bagian 5.