

Abstrak

Sistem pengenalan penutur menjadi salah satu bagian dari sistem pengenalan suara yang bertugas mengenali identitas seseorang berdasar isyarat tutur. Para peneliti dalam bidang tersebut melakukan analisis terhadap isyarat suara menggunakan berbagai metode ekstraksi ciri. Salah satu metode ekstraksi ciri yang populer digunakan adalah metode *Mel Frequency Cepstral Coefficient* (MFCC). Metode tersebut menghasilkan informasi ciri hanya dalam bentuk koefisien energi isyarat. Dari penggunaan metode MFCC saja tidak dimungkinkan untuk bisa mendapatkan bentuk ciri lain yang dapat mewakili informasi penutur. Di sisi lain, penggunaan spektrogram memungkinkan untuk diperolehnya bentuk ciri selain dalam bentuk koefisien.

Penelitian ini mengembangkan sistem pengenalan penutur dengan pengayaan ekstraksi ciri, yang tidak hanya memanfaatkan energi isyarat dalam bentuk koefisien untuk mengenali identitas penutur namun juga berdasar sejumlah informasi lain yang terkandung dalam isyarat tutur. Informasi tersebut dapat berupa koefisien serta citra visual spektrogram.

Metode yang diusulkan berupa representasi spektrogram yang memanfaatkan metode MFCC. Isyarat tutur yang menjadi masukan sistem, diolah sebelum masuk ke proses ekstraksi ciri. Dalam proses ekstraksi ciri, isyarat tutur dibagi menjadi sejumlah bingkai untuk kemudian diubah ke domain frekuensi. Dalam domain frekuensi, isyarat tutur diproses oleh bank tapis. Selanjutnya, isyarat yang dihasilkan dari penapisan diubah ke dalam skala mel untuk diubah lagi ke dalam bentuk desibel. Proses pengubahan ke desibel ini berbeda dari MFCC asli yang menggunakan proses DCT (*Discrete Cosine Transform*). Dari bentuk desibel tersebut, diekstrak sejumlah ciri dalam bentuk koefisien dan citra. Pengayaan ekstraksi ciri menggunakan representasi spektrogram yang memanfaatkan teknik MFCC belum banyak dikerjakan. Dataset suara yang digunakan dalam metode yang dikembangkan berupa rekaman dalam bahasa tutur Indonesia dengan total isyarat suara 12.600 detik yang dituturkan oleh 30 partisipan.

Penelitian ini menunjukkan dengan metode yang dikembangkan, diperoleh akurasi maksimal sebesar 96,35 %. Hasil ini lebih baik sebesar 6,15 % dibandingkan dengan representasi spektrogram serta lebih baik sebesar 6,98 % dibandingkan dengan metode MFCC tanpa representasi spektrogram.

Kata kunci: Spektrogram, MFCC, Ciri spektrogram.

Abstract

Speaker recognition system is a part of the sound recognition system whose recognize a person's identity based on the speech signals. Researchers in this area analyze sound signals using various methods of feature extraction. The most popular feature extraction method is Mel Frequency Cepstral Coefficient (MFCC) method. This method produces feature information only by the form of the signal energy coefficient. It is not possible for MFCC only to obtain other characteristics that can represent speakers information. On the other hand, the use of spectrograms makes it possible to obtain other characteristics than in the form of coefficients.

This study develops a speaker recognition system with enrichment feature, which not only utilizes signal energy in the form of coefficients to recognize the identity of the speaker but also based on a number of other information contained in the speech signal. This information can be in the form of coefficient and visual spectrogram image.

This study proposed a method based on spectrogram representation that utilizes the MFCC method. Voice cues that become the input of the system, are processed before entering the extraction process of features. In the feature extraction process, the voice signal is divided into a number of frames and then converted to the frequency domain. In the frequency domain, the voice signal is processed by the filter bank. Furthermore, the cues produced from are converted into the Mel's scale to be converted again into the form of decibels. The process of changing to this decibel's is different from the original MFCC that uses the DCT (Discrete Cosine Transform) process. From the form of decibels, a number of characteristics are extracted, i.e. in the form of coefficient and image. The feature extraction enrichment using spectrogram representations that utilize MFCC techniques have not been done much yet. The voice dataset used in the method is recorded in Indonesian language with a total of 12,600 seconds of voices spoken by 30 participants.

This study shows that the developed system obtain a maximum accuracy of 96.35 %. This result is better at 6.15 % compared to the spectrogram representation and better at 6.98 % compared to the MFCC method without spectrogram representation.

Keywords: *spectrogram, MFCC, spectrogram feature.*