

Representasi nilai hos dan model MFCC sebagai ekstraksi ciri pada sistem indentifikasi pembicara di lingkungan ber-noise menggunakan HMM

Agus Buono, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20426547&lokasi=lokal>

Abstrak

Suara merupakan suatu besaran yang memenuhi syarat sebagai ciri biometrik yang efektif dan efisien. Namun demikian, suara adalah fenomena yang merupakan perpaduan multidimensi serta dipengaruhi berbagai aspek, seperti karakteristik pembicara (dimensi titik artikularis, emosi, kesehatan, umur, jenis kelamin, dialek), bahasa, dan lingkungan (background dan media transmisi), sehingga sistem yang telah dikembangkan hingga sekarang belum bisa bekerja dengan baik pada situasi real. Hal inilah yang melatarbelakangi penelitian ini dilakukan.

Pada penelitian ini dilakukan kajian terhadap teknik higher order statistics (HOS) dan model Mel-Frequency Cepstrum Coefficients (MFCC) sebagai ekstraksi ciri yang diintegrasikan dengan Hidden Markov Model (HMM) sebagai pengenalan pola untuk menghasilkan sistem identifikasi pembicara yang lebih robust terhadap noise, khususnya Gaussian Noise. Penelitian yang dilakukan lebih difokuskan pada bagian ekstraksi Ciri dari sistem identifikasi pembicara. Sementara ini, bagian pengenalan pola menggunakan teknik yang telah banyak dikaji pada berbagai riset pemrosesan suara dan memberikan hasil yang baik, yaitu HMM. Strategi yang dilakukan adalah melalui pendekatan empiris untuk menunjukkan kegagalan teknik ekstraksi ciri konvensional, yaitu ID-MFCC yang berbasis power spektrum, pada lingkungan ber-noise, dilanjutkan dengan mengkaji permasalahannya, dan diusulkan teknik ekstraksi berbasis HOS untuk mengatasi permasalahan tersebut. Berikutnya adalah melakukan serangkaian percobaan untuk menunjukkan efektifitas teknik yang diusulkan, studi komparasi dan mengajukan suatu usulan rancangan sistem.

Berdasar bukti empiris, terlihat bahwa permasalahan 1D-MFCC adalah pada inputnya, yaitu power spektrum yang bersifat tidak stabil terhadap noise. Pada penelitian ini diusulkan untuk mengganti power spektrum dengan bispektrum yang secara teori lebih robust terhadap noise. Teknik yang diusulkan adalah suatu metodologi untuk mengekstrak nilai bispektrum sinyal suara dengan MFCC dan diintegrasikan dengan HMM untuk membentuk sistem identifikasi pembicara. Oleh karena itu, pada penelitian ini dilakukan perluasan teknik ID-MFCC menjadi 2D-MFCC. Untuk meningkatkan efektifitas sistem, diusulkan teknik kuantisasi sebagai cara merepresentasikan nilai bispektrum sehingga distribusi spasialnya terakomodasi, dan dilanjutkan dengan transformasi wrapping dan kosinus seperti pada MFCC.

Hasil percobaan menunjukkan bahwa teknik konvensional yang berbasis pada power spektrum dapat menangkap ciri suara tanpa penambahan noise dengan baik dan jika dipadukan dengan Mel-Frequency Cepstrum Coefficients (MFCC) sebagai ekstraksi ciri dan HMM sebagai pengenalan pola, maka akan menghasilkan sistem dengan akurasi di atas 98.8%. Namun demikian, dengan penambahan noise 20 dB, nilai power spektrum mengalami perubahan secara nyata, sehingga akurasi sistem jatuh hingga level di bawah 50%. Teknik penghapusan noise secara adoptive mampu meningkatkan akurasi menjadi 77.7%, namun dengan noise yang lebih besar, teknik ini gagal bekerja dengan baik.

Sistem yang dikembangkan dengan menggunakan bispektrum sebagai penentu ciri dipadukan dengan MFCC yang diperluas ke dua dimensi berhasil memberikan akurasi 99.9% untuk sinyal suara asli. Namun

untuk sinyal dengan noise 20 dB, akurasi sistem menjadi sekitar 70%. Optimasi pada bentuk Elter pada proses MFCC dengan algoritma genetika mampu meningkatkan akurasi menjadi 88.8%. Akan tetapi dengan noise yang lebih tinggi, sistem gagal bekerja dengan baik.

Teknik kuantisasi skalar terhadap nilai bispektrum yang dilanjutkan dengan proses wrapping dan transformasi kosinus seperti yang dilakukan pada MFCC mampu meningkatkan robustness sistem terhadap noise dengan akurasi 99.5% dan 83% masing-masing untuk sinyal asli dan sinyal dengan penambahan noise 20 dB. Namun untuk noise 10 dB, teknik ini gagal bekerja dengan baik. Dari percobaan dengan teknik kuantisasi vektor, terlihat bahwa rata-rata nilai bispektrum di atas kuartil tiga adalah penduga terbaik bagi nilai bispektrum setiap channel dengan jumlah 400 channel. Selain itu nilai parameter yang optimum pada proses ekstraksi ciri dengan kuantisasi vektor dilanjutkan dengan proses wrapping dan transformasi kosinus pada sinyal dengan penambahan noise adalah jarak filter linear 75, jarak Elter logaritma 1.06 dan proporsi filter linear dan logaritma 30:20. Kinerja sistem menunjukkan peningkatan yang berarti dengan akurasi 88% dan 75.5% masing-masing untuk sinyal dengan penambahan noise 20 dB dan 10 dB. Namun demikian untuk sinyal asli justru lebih rendah, yaitu dengan akurasi maksimum hanya 94.5%. Hal ini berarti bahwa teknik ekstraksi ciri yang efektif tergantung dari kualitas sinyal masukan. Oleh karena itu sistem yang dikembangkan sebaiknya dilengkapi di bagian awalnya dengan kemampuan untuk menduga kualitas sinyal masukan.

Dari studi eksploratif terhadap nilai autokorelasi dan ragam sinyal suara, diperoleh bahwa kualitas sinyal dapat diidentifikasi dengan besaran yang dirumuskan sebagai negatif dari logaritma perkalian nilai absolut autokorelasi dari lag 1 hingga lag 21. Nilai ambang untuk membedakan sinyal sesuai kualitasnya dengan besaran tersebut adalah di antara 7 hingga 15. Jika nilai besaran tersebut kecil, maka teknik 1D-MFCC lebih sesuai untuk diterapkan. Sedangkan untuk hal lainnya, disarankan menggunakan teknik kuantisasi vektor terhadap nilai bispektrum sebagai pengekstraksi ciri. Berdasar nilai ambang inilah disusun prototipe sistem identifikasi pembicara menggunakan software Matlab.

.....Mel-Frequency Cepstrum Coefficients (MFCC) as speech signal feature extraction technique and integrated with Hidden Markov Model (HMM) as classifier to form a speaker identification system that more robust to Gaussian Noise. The experiments is focused on the subsystem of feature extraction, whereas in the subsystem of classifier, we use the HMM. In this research, we show the ineffectiveness of 1D-MFCC as feature extraction in the noisy environment empirically, analysis the problem and propose some techniques for feature extraction to handle the problem. Next, we conduct a series of experiments to show the effectiveness of the propose methods. Finally, we make a comparison among methods to capture the characteristics of each and propose a prototype of speaker identification system.

According to the result, the main problem with 1D-MFCC is in the aspect of its input, i.e. power spectrum. This quantity is not stable enough with existing noise. In this research we replace the power spectrum by bispectrum that more robust to noise. Then, the propose methods is focused on how to extract the bispectrum value and integrate with HMM to form the speaker identification system. Firstly, 1D-MFCC extended into 2D-MFCC, so the technique workable for bispectrum value as the input. In order to improve the system performance, we use scalar and vector quantization for bispectrum value representation and continue with wrapping and cosines transform prior to classifier process.

The experiments show that the conventional method based on power spectrum (1D-MFCC) gives a good result for signal without addition by Gaussian noise, with 98.8% of accuracy. Nevertheless, with noise only 20 dB, the system performance drop significantly with accuracy below 50%. The noise canceling technique

can improve the accuracy up to 77.7%, but fails for noise more than 20 dB. The 2D-MFCC that developed using bispectrum as speech signal feature gives 99.9% of accuracy for original signal and 88.8% for signal corrupted by 20 dB of noise. Compare with ID-MFCC, this system performance is higher. Nevertheless, for noise more than 20 dB, the system fails.

In order to improve the system performance, we propose scalar and vector quantization for representation the bispectrum value, and continue with wrapping and cosines transform prior to classifier process. The vector quantization technique yield the system more stable with noise, and gives the highest recognition compare with others, especially for signal corrupted by noise. The accuracy for signal with addition by 20 dB and 10 dB of noise are 89% and 75.5%, respectively. But, for original signal, the accuracy is only around 90%. It means the effective technique for feature extraction depend on the quality of input signal.

According to the exploration of autocorrelation of speech signal, it is shown that the signal quality can be divided by the negative value of multiplication of absolute value of its autocorrelation from lag 1 until lag 21. The threshold lies between 7 and 15. If the value is small enough, it is better for use the ID-MFCC technique. Otherwise, we advise to use the system based on bispectrum represented by vector quantization and continue by the wrapping and cosines transform prior to the classifier process. By using this threshold, we propose a prototype for speaker identification system developed by Matlab software.