JUPITER
Teknologi Informatika & Komputer
Vol. 3 No. 1 Februari 2022
Institut Bisnis Muhammadiyah Bekasi
P-ISSN: 2798-3129, E-ISSN: 2798-6187

PEMODELAN SISTEM IDENTIFIKASI PEMBICARA DENGAN MFCC DAN SUPPORT VECTOR MACHINE

Luthfan Almanfaluthi[1], Agus Buono[2], Yani Nurhadryani[3]

[1]Sekolah Tinggi Bahasa Asing JIA, luthfan.a@stba-jia.ac.id
[2]Institut Pertanian Bogor, pudesha@yahoo.co.id
[3]Institut Pertanian Bogor, yhadryani@yahoo.com

ABSTRACT

In this paper, we focus on speech recognition were speakers used text-dependent which means the text agreed in advance and will be used next. This system using MFCC as feature extraction and SVM as pattern recognition. Data were taken from 10 adult speakers with differences in gender, age and ethnicity. Each speakers provide 50 ballot "computer" and its pronunciation is not controlled resulting in 500 data. Some data training are contaminate with gaussian noise with level 80dB, 70dB, 60dB, 50dB, 40dB, 30dB, 20dB, 10dB and 0dB. The research uses a frame length: 40 ms, overlapping frames: 50%, and the coefficient mel: 13. Noise Cancelling also tested in this research, although not getting optimal results. Pattern recognition SVM with RBF kernel functions produce 100% accurate results. Time process of Sequential Minimal Optimization algorithm is better than Quadratic Programming algorithms. Increasing the number of speakers to see the performance of the system with a greater amount of data can be made for further research.

Keywords: *Speaker Identification, MFCC, SVM, Noise Cancelling*

ABSTRAK

Dalam artikel ini, kita fokus pada pengenalan suara dimana pembicara menggunakan teks yang diatur yang berarti teks disetujui terlebih dahulu dan akan digunakan selanjutnya. Sistem ini menggunakan MFCC sebagai ekstraksi ciri dan SVM sebagai pengenalan pola. Data diambil dari 10 pembicara dewasa dengan perbedaan jenis kelamin, usia dan suku. Setiap pembicara menghasilkan 50 suara "komputer" dan pengucapannya tidak terkontrol sehingga menghasilkan 500 data. Beberapa data training terkontaminasi gaussian noise dengan level 80dB, 70dB, 60dB, 50dB, 40dB, 30dB, 20dB, 10dB dan 0dB. Penelitian menggunakan panjang frame: 40 ms, overlapping frame: 50%, dan koefisien mel: 13. Noise Cancelling juga diuji dalam penelitian ini, meskipun belum mendapatkan hasil yang optimal. SVM sebagai pengenalan pola dengan fungsi kernel RBF menghasilkan hasil yang akurat 100%. Proses waktu algoritma Sequential Minimal Optimization lebih baik daripada algoritma Quadratic Programming. Penambahan jumlah pembicara untuk melihat kinerja sistem dengan jumlah data yang lebih banyak dapat dilakukan untuk penelitian selanjutnya.

Kata Kunci: Identifikasi Pembicara, MFCC, SVM, Noise Cancelling

PENDAHULUAN

Setiap hari manusia bertukar informasi dengan menggunakan media suara walaupun dapat juga bertukar informasi dengan media teks dan alat bantu semacamnya. Sinyal suara yang diucapkan setiap manusia memiliki karakter dan kualitas yang berbeda atau bersifat unik. Sinyal suara dipengaruhi banyak hal, seperti intra-speaker variability (dimensi artikularis pembicara, emosi, kesehatan, umur, jenis kelamin, dialek) dan noise (latar belakang suara lingkungan dan media transmisi).

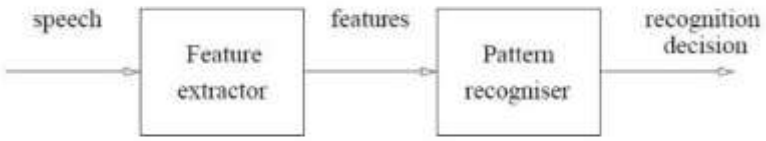
Suara dapat juga dikategorikan sebagai alat biometrik karena memiliki ciri-ciri sebagai berikut: alami, mudah diukur, tidak terlalu berubah seiring waktu atau kondisi fisik, tidak terlalu terganggu dengan adanya gangguan lingkungan, serta tidak mudah ditiru. Suara hampir memenuhi semua persyaratan biometrik, namun permasalahan yang timbul dari pemrosesan suara yaitu suara adalah bersifat multidimensi (linguistik, semantik, artikularis dan akustik).

Proses identifikasi dengan suara memiliki keuntungan secara ekonomis dibandingkan dengan identifikasi secara biometrik lainnya seperti identifikasi pada wajah, sidik jari, tanda tangan, retina dan lain-lain.

Identifikasi dengan suara hanya membutuhkan alat tambahan berupa mikrofon dan kartu suara, sedangkan karakteristik lain membutuhkan alat tambahan seperti scanner. Hal ini dapat menekan sedikit biaya pengembangan sistem.

Sinyal suara manusia mempunyai tingkat variabilitas yang sangat tinggi. Suatu sinyal suara yang dikeluarkan oleh pembicara yang berbeda-beda menghasilkan pola ucapan yang berbeda-beda pula. Masyarakat Indonesia mempunyai beragam suku dan budaya, sehingga banyak permasalahan pola ucapan yang berbeda-beda untuk satu kata yang sama. Oleh karena permasalahan ini bisa menjadi problem dalam sistem identifikasi pembicara, sehingga perlu dikembangkan suatu sistem yang relatif lebih robust terhadap permasalahan intra-speaker variability dan noise. Sistem identifikasi pembicara lebih berfokus pada analisis dengan dua subsistem yaitu Feature Extractor (ekstraksi ciri) dan Pattern Recogniser (pengenalan pola) yang diilustrasikan oleh Gambar 1.

Gambar 1. Sistem Identifikasi Pembicara



Penelitian tentang sistem identifikasi pembicara telah banyak dilakukan oleh para peneliti, seperti: penelitian tentang identifikasi pembicara menggunakan ekstraksi fitur MFCC telah dilakukan dan menghasilkan tingkat akurasi mendekati 100%. Penelitian ini menggunakan 1D-MFCC mendapatkan akurasi 98,8%, sedangkan 2D-MFCC mendapatkan hasil akurasi 99,9% pada sinyal suara tanpa noise. Sedangkan untuk SVM pengenalan pola pada sinyal suara telah dilakukan dengan baik dan mendapatkan hasil yang menakjubkan. Identifikasi pembicara ini menggunakan SVM dengan sumber 20 laki-laki dan 20 perempuan dari database Aurora-2. Mereka mengujinya tanpa noise pada level 8000 Hz dan menghasilkan akurasi 95,1%.

TINJAUAN PUSTAKA

Prinsip Identifikasi Pembicara

Identifikasi pembicara adalah proses mengklasifikasikan pembicara dari sejumlah suara pembicara yang diberikan, sebagai suatu keputusan yang terbaik. Dasar kerja sistem identifikasi pembicara yaitu mampu meniru kemampuan manusia dalam mengenal identitas seseorang melalui suara yang didengar, sehingga sistem identifikasi pembicara dapat dimasukan kedalam kelompok sistem kecerdasan buatan.

Secara garis besar terdapat dua tahap proses yang dilibatkan untuk membangun suatu sistem identifikasi pembicara. Pertama, mendapatkan informasi spesifik atau nilai ciri dari suara yang diamati. Kedua, mengklasifikasikan suara melalui proses pencocokan nilai ciri suara yang diterima dengan nilai ciri suara acuan (basis data ciri suara).

Dari sudut pandang linguistik, terdapat dua metode yang dapat diterapkan untuk mengembangkan sistem identifikasi pembicara. Metode pertama disebut text-dependent, dan metode kedua disebut text-independent.

Mel-Frequency Cepstrum Coefficients (MFCC)

Ekstraksi ciri adalah proses untuk menentukan vektor yang dapat digunakan sebagai penciri objek atau individu. Ciri yang biasa digunakan adalah koefisien cepstral. MFCC merupakan ekstraksi ciri yang menghitung koefisien cepstral dengan mempertimbangkan pendengaran manusia. MFCC memiliki tahapan yang terdiri atas:

a. **Frame Blocking**. Pada tahap ini sinyal suara continous speech dibagi ke dalam beberapa frame serta dilakukan overlapping frame agar tidak kehilangan informasi.
b. **Windowing**. Merupakan salah satu jenis filtering untuk meminimalisasikan distorsi antar frame. Proses ini dilakukan dengan mengalikan antar frame dengan jenis window yang digunakan.
c. **Fast Fourier Transform (FFT)**. Tahapan selanjutnya adalah mengubah tiap frame dari domain waktu ke dalam domain frekuensi. FFT adalah algoritme yang mengimplementasikan Discrete Fouries Transform (DFT).
d. **Mel-Frequency Wrapping**. Persepsi sistem pendengaran manusia terhadap frekuensi sinyal suara ternyata tidak hanya bersifat linear. Penerimaan sinyal suara untuk frekuensi rendah (< 1000 Hz) bersifat linear, sedangkan untuk frekuensi tinggi (> 1000 Hz) bersifat logaritmik.
e. **Cepstrum**. Tahap ini merupakan tahap terakhir MFCC. Pada tahap ini mel-frequency akan diubah menjadi domain waktu menggunakan Discrete Cosine Transform (DCT).

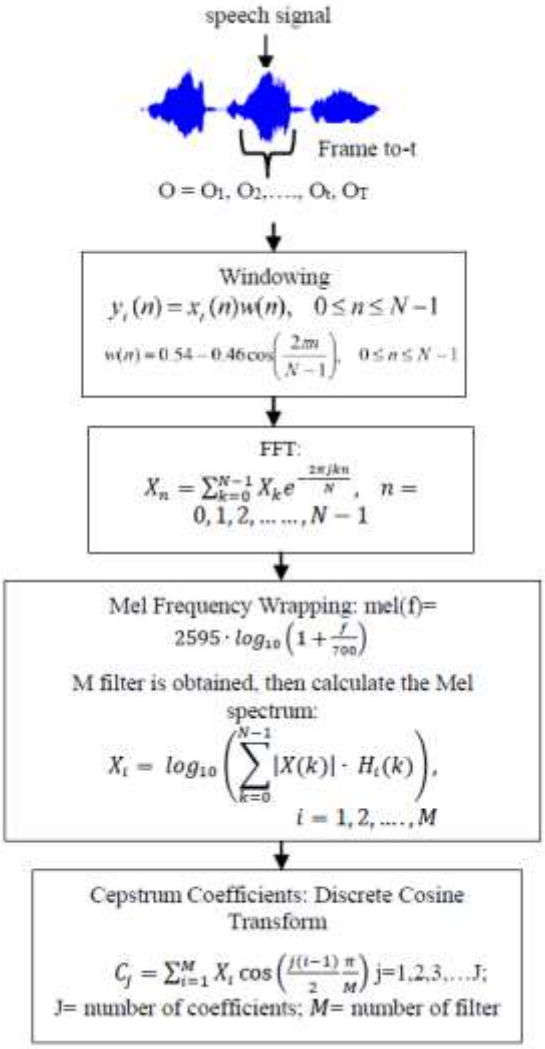
Proses MFCC dapat dilihat pada Gambar 2.

Support Vector Machine (SVM)

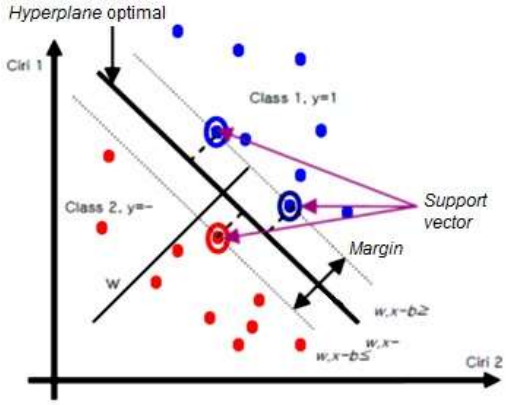
SVM adalah salah satu teknik klasifikasi data dengan proses pelatihan (supervised learning). Salah satu ciri dari metode klasifikasi SVM adalah menemukan garis pemisah (hyperplane) terbaik sehingga diperoleh ukuran margin yang maksimal. Margin adalah jarak antara hyperplane tersebut dengan titik terdekat dari masing-masing kelas. Titik yang paling dekat ini disebut dengan support vector. Untuk penelitian ini digunakan SVM dengan algoritma Sequential Minimal Optimization (SMO) dan algoritma Quadratic Programming (QP).Ilustrasi SVM untuk linear separable data dapat dilihat pada Gambar 3.

JUPITER
Teknologi Informatika & Komputer
Vol. 3 No. 1 Februari 2022
Institut Bisnis Muhammadiyah Bekasi
P-ISSN: 2798-3129, E-ISSN: 2798-6187

Gambar 2. Tahapan Mel-Frequency Cepstrum Coefficients (MFCC)



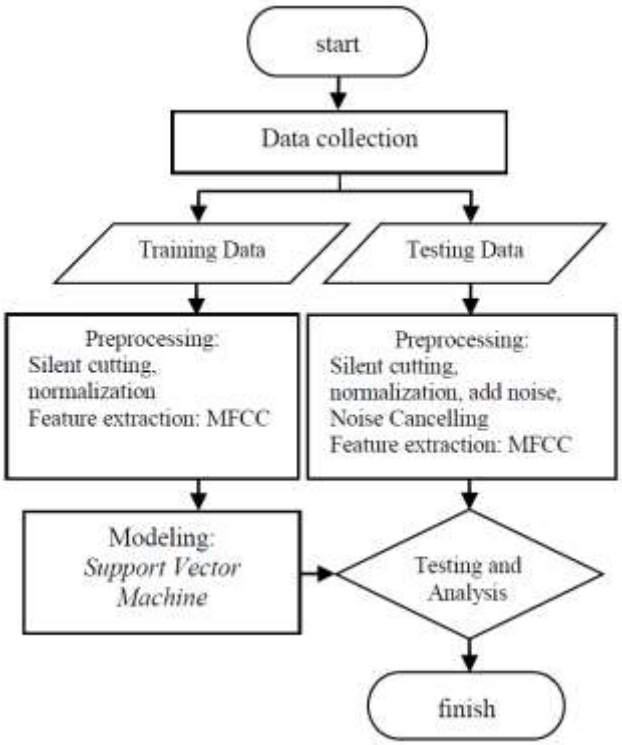
Gambar 3. SVM dengan data terpisah secara linear



METODE PENELITIAN

Metodologi penelitian ini terdiri dari beberapa tahapan proses yaitu pengumpulan data, preprocessing, pemodelan SVM, pengujian dan analisis, dan pembuatan interface. Metodologi proses sistem identifikasi pembicara ditunjukkan pada Gambar 4.

Gambar 4. Metodologi Sistem Identifikasi Pembicara



HASIL DAN PEMBAHASAN

Pengambilan data suara dilakukan dengan merekam suara menggunakan alat mikrofon. Sumber suara diperoleh dari 10 orang pembicara dewasa dengan perbedaan jenis kelamin, umur dan suku yang masing-masing mengucapkan 50 kali kata “KOMPUTER” yang pengucapannya tidak dikontrol hingga didapatkan 500 data suara. Durasi rekam yang digunakan yaitu 2 detik dengan besar frekuensi rekam 16KHz dan data suara disimpan dalam format audio dengan ekstensi (*.wav).

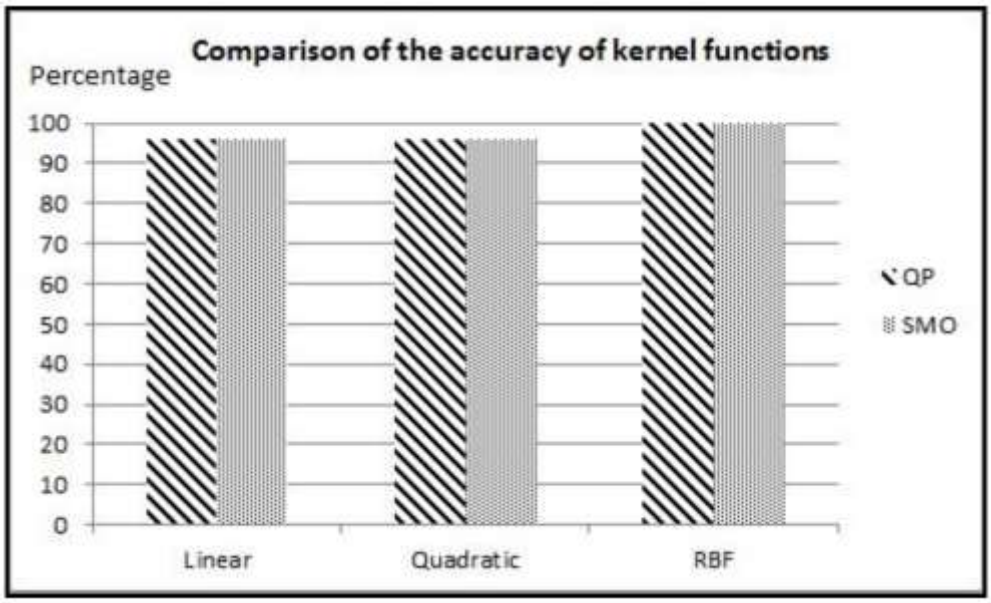
Dalam hal intra-speaker variability (jenis kelamin, umur dan suku) maka pada tahap pengambilan data suara dari 10 orang pembicara didapatkan rentang umur yang beragam yaitu dari umur paling rendah 16 tahun dan paling tinggi umur 42 tahun. Sedangkan untuk jenis kelamin didapatkan lima orang berjenis kelamin wanita dan lima orang berjenis kelamin pria. Untuk perbedaan suku, didapatkan tiga suku yang berbeda yaitu empat orang bersuku sunda, lima orang dari suku jawa dan satu orang dari suku betawi. Karakteristik kesepuluh pembicara tersebut disajikan pada Tabel 1.

Tabel 1. Daftar 10 pembicara yang digunakan dalam penelitian

Pembicara	Jenis Kelamin	Usia (tahun)	Suku
1	Wanita	30	Sunda
2	Wanita	31	Jawa
3	Wanita	16	Jawa
4	Pria	25	Jawa
5	Pria	19	Sunda
6	Pria	41	Jawa
7	Wanita	33	Betawi
8	Wanita	42	Jawa
9	Wanita	22	Sunda
10	Pria	28	Sunda

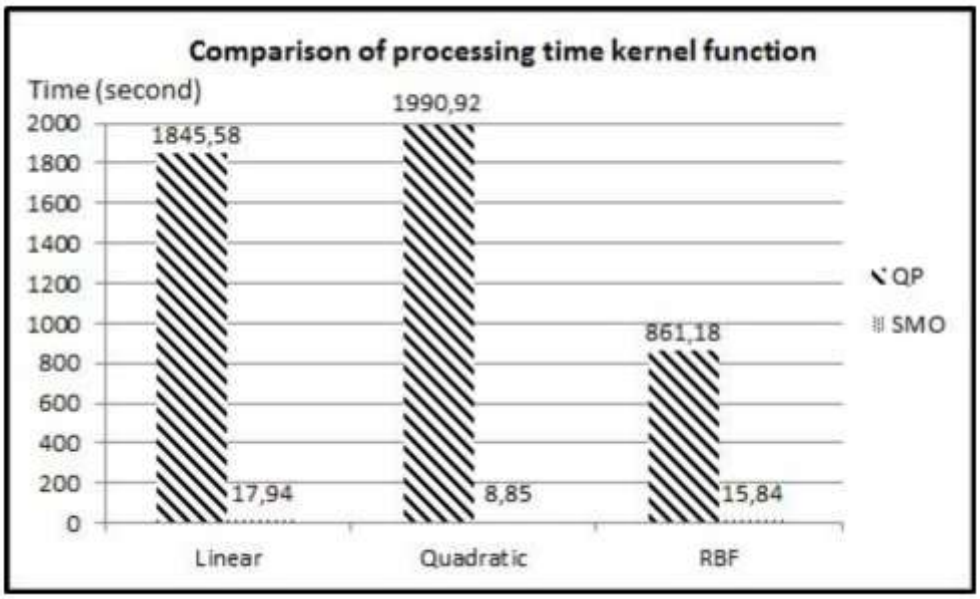
Penelitian ini menggunakan software Matlab R2010b versi 7.11.0.584. Hasil perbandingan akurasi dengan fungsi kernel SVM ditunjukkan pada Gambar 5. Dari gambar tersebut fungsi kernel RBF menghasilkan akurasi 100%, sedangkan untuk fungsi lainnya masih terdapat kesalahan nilai yang dihasilkan sebesar 96%. Algoritma difference juga diuji dan menghasilkan hasil yang serupa antara algoritma SMO dan algoritma QP.

Gambar 5. Perbandingan Akurasi Fungsi Kernel



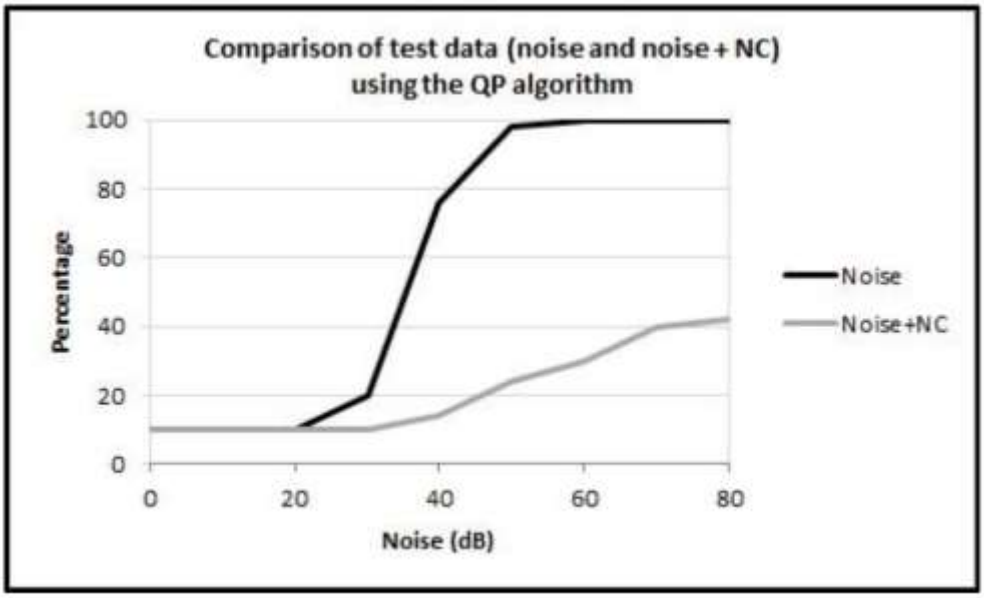
Waktu proses algoritma SMO lebih cepat dibandingkan dengan algoritma QP, hasilnya dapat dilihat pada Gambar 6.

Gambar 6. Perbandingan waktu pemrosesan fungsi kernel

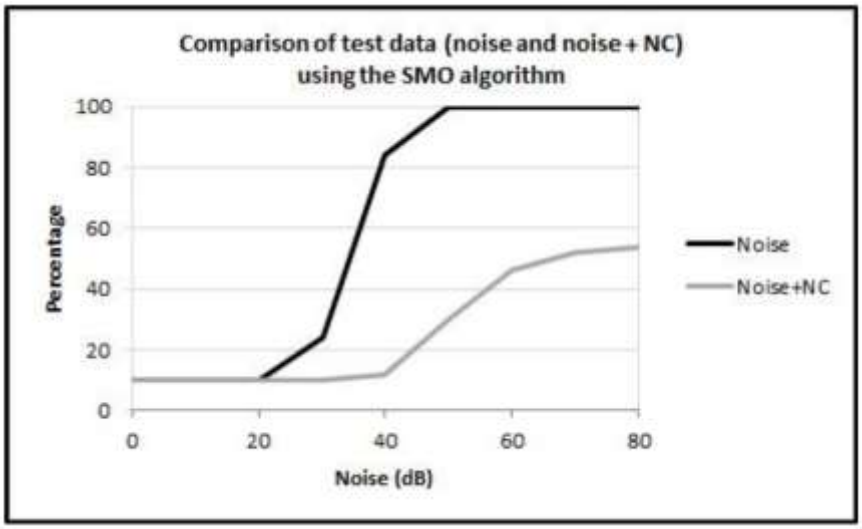


Hasil percobaan menggunakan tambahan noise, dapat dilihat pada Gambar 7 dan Gambar 8. Pada Gambar 7 merupakan hasil percobaan menggunakan algoritma QP dan pada Gambar 8 merupakan hasil percobaan menggunakan algoritma SMO. Kedua gambar tersebut dapat dilihat, bahwa sistem dapat bertahan tanpa error hanya dengan tambahan noise 50db sehingga untuk akurasi dibawah noise 40db terlihat jelas.

Gambar 7. Perbandingan data pengujian (Noise dan Noise + NC) menggunakan algoritma QP



Gambar 8. Perbandingan data pengujian (Noise dan Noise + NC) menggunakan algoritma SMO



PENUTUP

Simpulan

Pengenalan pola SVM dengan fungsi kernel RBF menghasilkan hasil yang akurat 100%. Noise Cancelling juga diuji dalam penelitian ini, walaupun tidak mendapatkan hasil yang optimal Waktu proses algoritma Sequential Minimal Optimization (SMO) lebih baik daripada algoritma Quadratic Programming (QP).

Saran

Penambahan jumlah speaker untuk melihat kinerja sistem dengan jumlah data yang lebih banyak dapat dilakukan untuk penelitian selanjutnya.

REFERENSI

Reynolds D. 2002. Automatic Speaker recognition Acoustics and Beyond. Tutorial note, MIT Lincoln Laboratory.

Campbell JP. 1997. Speaker Recognition: A Tutorial. Proceedings of the IEEE Vol. 85 No. 9.

Srinivasamurthy N. 2006. Compression Algorithms for Distributed Classification with Applications to Distributed Speech Recognition. A Dissertation Presented to the Faculty Of The Graduate School, University Of Southern California.

Guiwen O and Dengfeng K. 2004. Text- independent speaker verification based on relation of MFCC components. 2004 International Symposium on Chinese Spoken Language Processing, pp. 57-60.

Mezghani A and O'Shaughnessy D. 2005. Speaker verification using a new representation based on a combination of MFCC and formants. 2005 Canadian Conference on Electrical Ana Computer Engineering, pp. 1461-1464.

Homayounpour M and Rezaian I. 2008. Robust Speaker Verification Based on Multi Stage Vector Quantization of MFCC Parameters on Narrow Bandwidth Channels, ICACT 2008, vol 1, pp.336-340.

Lin CC, Chen SH, Truong TK, and Chang Y. 2005. Audio Classification and Categorization Based on Wavelets and Support Vector Machine, IEEE Trans. on Speech Ana Audio Processing, Vol. 13, No. 5, pp. 644-651.

Buono A. 2009. Representasi Nilai HOS dan Model MFCC sebagai Ekstraksi Ciri pada Sistem Identifikasi Pembicara di Lingkungan Bernoise Menggunakan HMM. [dissertation]. Depok: Computer Science Department, University of Indonesia.

Chen S and Luo Y. 2009. Speaker Verification Using MFCC and Support Vector Machine. Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I, Hong Kong.

Mak G. 2000. The Implementation of Support Vector Machine Using The Sequential Minimal Optimization Algorithm. Master Degre. McGill University.