PENGENALAN SUARA MENGGUNAKAN MEL FREQUENCY CEPSTRAL COEFFICIENTS DAN SELF ORGANIZING MAPS

Meutia Puspitasari[1], Julian Supardi, M.T.[2], Yoppy Sazaki, M.T.[3]

[1,2,3]Fakultas Ilmu Komputer Universitas Sriwijaya, Palembang

[1]meutiaps@gmail.com [2 3]yoppysazaki@gmaill.com

ABSTRACT

The interaction between human and computer can not be enjoyed by everyone, especially for people with disabilities, they will be difficult to interact using a computer. Speech to text recognition is one way to make it easy for anyone to operate a computer. Therefore, developed speech to text recognition research using Mel Frequency Cepstral Coefficients (MFCC) method to extract features from the voice data and Self Organizing Maps (SOM) to classify voice. This study uses primary data of 200 voice data as input for training and testing steps. The accuracy of the results achieved in this study is, testing for the same data with training data is 100 %, while for the data that is different from training data is 82 %. The results is obtained by comparing the amount of voice data that recognized successfully by the entire amount of voice data that tested.

Keywords: Speech to Text Recognition, Mel Frequency Cepstral Coefficients, Self Organizing Maps

PENDAHULUAN

Pengenalan suara adalah salah satu cara untuk memudahkan bagi siapa saja untuk mengoperasikan komputer, terutama bagi orang-orang dengan kekurangan fisik. Pengenalan suara memiliki potensi besar untuk menjadi faktor penting dari interaksi antara manusia dan komputer. [20]

Mel Frequency Cepstrum Coefficients (MFCC) ekstraksi fitur secara luas digunakan dalam pengenalan suara. MFCC diperkenalkan oleh Davis dan Mermelstein pada tahun 1980 mengusulkan menggunakan MFCC ekstraksi fitur untuk sistem pengenalan suara. Algoritma MFCC mengurangi kekuatan komputasi sampai 53% dibandingkan dengan algoritma konvensional. Hasil simulasi menunjukkan algoritma MFCC memiliki akurasi 92,93 %. [6]

Self Organizing Maps (SOM) adalah model jaringan saraf tiruan dengan metode pembelajaran tidal terawasi. Salah satu keunggulan dari algoritma SOM adalah mampu memetakan data dalam dimensi tinggi ke bentuk peta rendah dimensi. Menurut penelitian sebelumnya penggunaan Kohonen atau SOM dalam pengenalan suara otomatis mendapatkan akurasi 91%.[12]

Dalam penelitian ini akan mengembangkan sistem untuk pengenalan suara yang berfokus pada angka dengan menerapkan MFCC dan SOM. Data primer adalah sebagai masukan untuk langkah pelatihan dan pengujian. Hasil penelitian ini akan mengkonversi sinyal suara manusia ke dalam bentuk teks.

METODE PENELITIAN

Pada penelitian ini akan dilakukan pengembangkan arsitektur jaringan syaraf tiruan SOM untuk pengenalan suara dengan metode ekstraksi fitur MFCC dan melakukan analisa dan pembahasan terhadap hasil pengujian perangkat lunak.

Tujuan penelitian ini adalah :

1. Mengembangkan perangkat lunak pengenalan suara menggunakan MFCC untuk mengekstrak fitur serta algoritma *clustering* SOM untuk mengenali sinyal suara.
2. Mengukur tingkat keakuratan metode MFCC serta SOM dalam mengenali suara yang diucapkan.

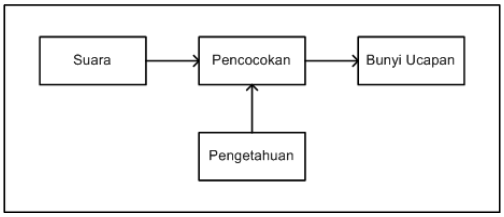
Dalam penelitian ini dirumuskan permasalahan yang akan dibahas yaitu apakah metode *Mel Frequency Cepstrum Coeffisiens* (MFCC) dan *Self Organizing Maps* (SOM) merupakan metode yang baik untuk sistem pengenalan suara.

Untuk lebih memokuskan pengerjaan penelitian ditetapkan pembatasan-pembatasan sebagai berikut:

1. Kata yang digunakan adalah kata dalam bahasa Indonesia untuk angka 1-10, yaitu “Satu”, “Dua”, “Tiga”, “Empat”, “Lima”, “Enam”, “Tujuh”, “Delapan”, “Sembilan”, dan “Sepuluh”.
2. Data yang digunakan merupakan data digital suara, mono, dan direkam dalam lingkungan yang tenang.
3. Data suara dalam format WAV.
4. Perangkat lunak yang dihasilkan hanya dapat mengenali suara yang telah ditetapkan. Output dari penelitian ini akan mengkonversikan sinyal suara manusia ke dalam teks.
5. Data suara yang digunakan berjumlah 20 data suara per satu kata dari 20 orang responden. Pada proses training digunakan 15 data suara dari setiap kata, sedangkan pada proses pengenalan digunakan 5 data suara dari setiap kata.
6. Data suara tidak secara *real time*.

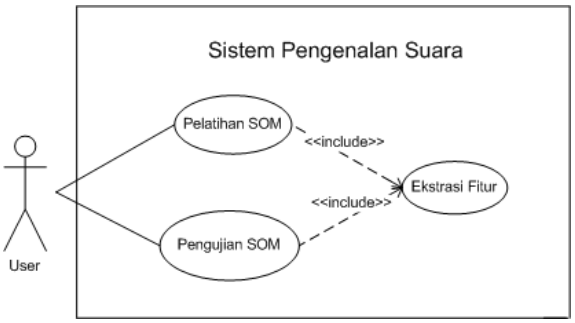
HASIL DAN PEMBAHASAN

Sistem yang dirancang pada penelitian ini adalah aplikasi pengenalan suara menggunakan metode *mel frequency cepstral coefficients* untuk mengekstraksi fitur dan *self organizing maps* untuk mengenali suara. Aplikasi ini dapat mengenali suara yang berfokus pada angka. Data primer adalah sebagai masukan untuk langkah pelatihan dan pengujian. Hasil penelitian ini akan mengkonversi sinyal suara manusia ke dalam bentuk teks.



Gambar 1. Mekanisme Pengenalan Suara

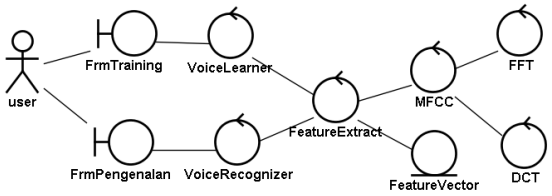
Sistem dirancang dengan menggunakan metode *Unified Modelling Language* (UML) Dalam metode ini digunakan pemodelan *use case*.



Gambar 2. *Use Case*

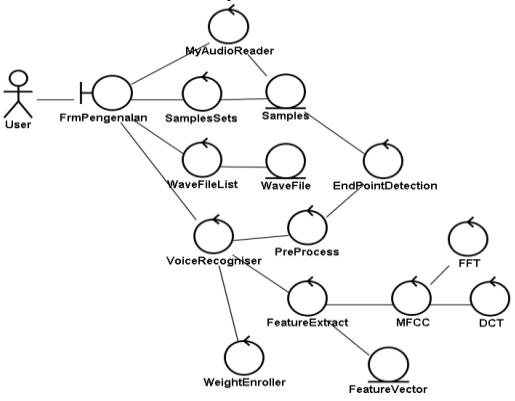
Gambar diagram di atas menggambarkan nama-nama *use case*, aktor, dan hubungannya dari sistem pengenalan suara. Untuk penjelasan rinci dari *use case* di atas dijelaskan pada sub-bab selanjutnya.

Pada bagian ini akan dibahas diagram kelas analisis dari *use case* yang telah didefinisikan. Gambar 3. merupakan diagram kelas analisis dari proses Ekstraksi Fitur.



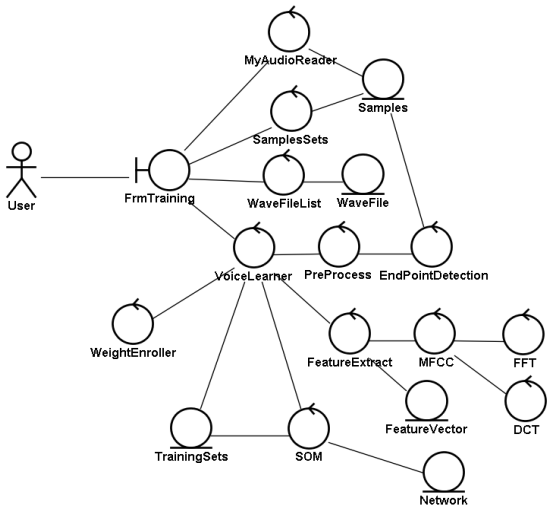
Gambar 3. Diagram Kelas Analisis Proses Ekstraksi Fitur

Pada Gambar 4. merupakan diagram kelas analisis dari proses Pelatihan SOM.



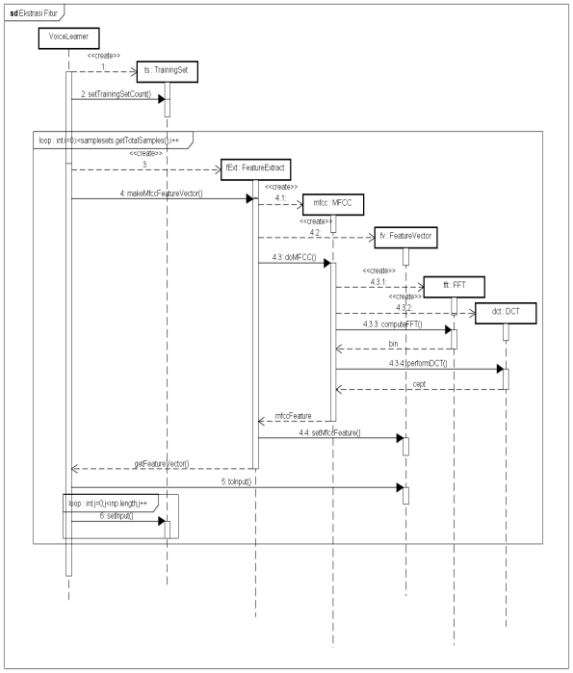
Gambar 4. Diagram Kelas Analisis Proses Pelatihan SOK

Gambar 5. merupakan diagram kelas analisis dari proses Pengujian SOM.



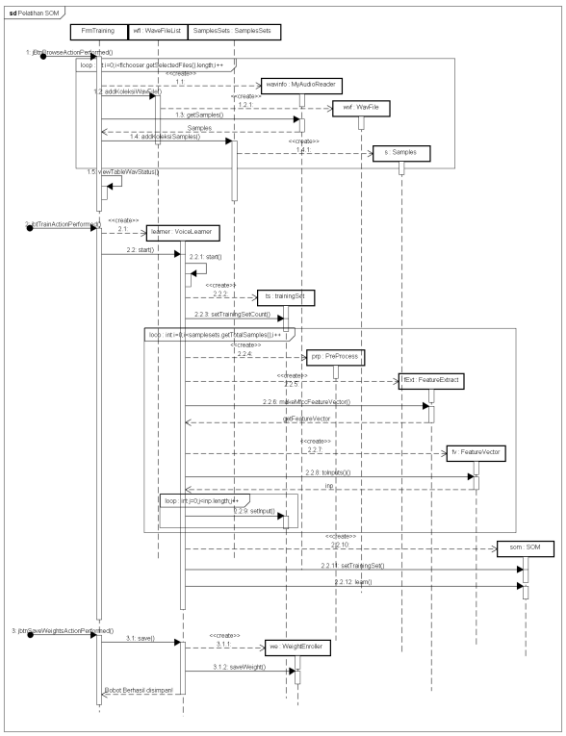
Gambar 5. Diagram Kelas Analisis Proses Pengujian SOM

Selanjutnya akan dibahas diagram *sequence* yang merupakan diagram yang menjelaskan pola interaksi antara objek-objek yang disusun dalam urutan kronologis. Diagram ini menunjukkan objek-objek yang berpartisipasi dalam interaksi dan pesan-pesan yang dikirimkan dari *use case* yang telah didefinisikan. Gambar 6. merupakan diagram *sequence* untuk *use case* Ekstraksi Fitur.



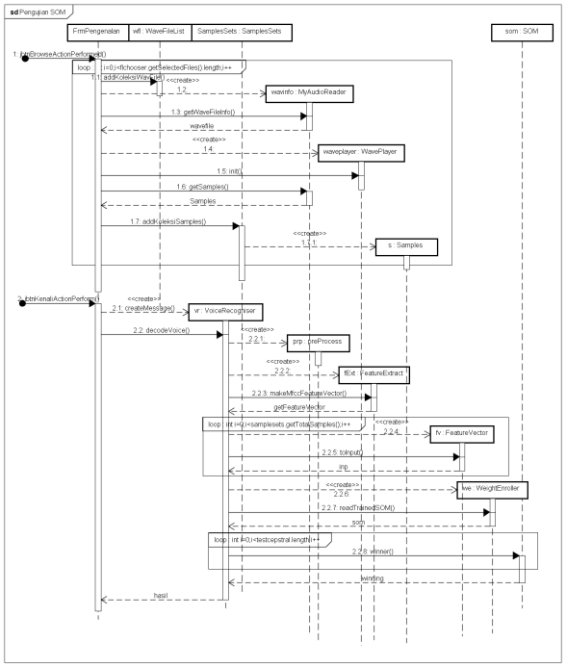
Gambar 6. Diagram *Sequence* dari Proses Ekstraksi Fitur

Gambar 7. merupakan diagram *sequence* dari proses Pelatihan SOM.



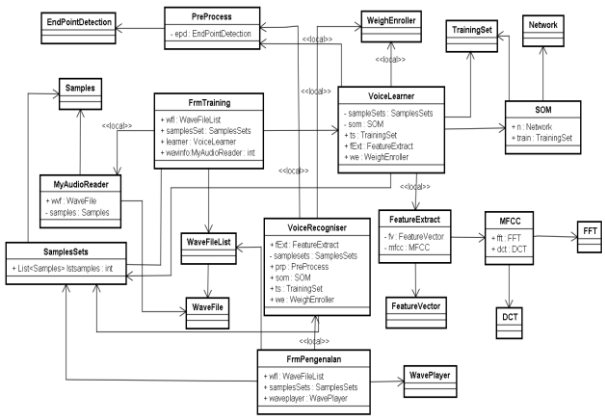
Gambar 7. Diagram *Sequence* dari Proses Pelatihan SOM

Gambar 8. merupakan diagram *sequence* dari proses Pengujian SOM.



Gambar 8. Diagram *Sequence* dari Proses Pengujian SOM

Selanjutnya pada Gambar 9. menjelaskan digram kelas keseluruhan dari sistem pengenalan suara.

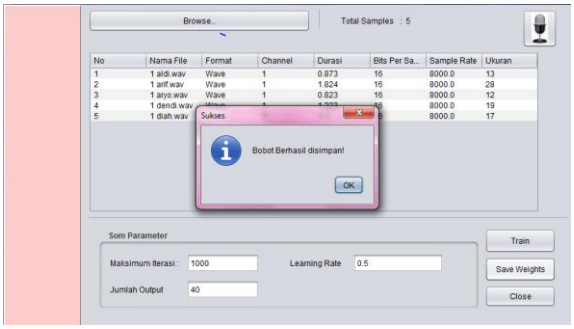


Gambar 9. Diagram Kelas Keselurahan Sistem Pengenalan Suara

Untuk mendapatkan data sebagai bahan analisa maka dibutuhkan suatu pengujian terhadap fungsional sistem dan pengujian terhadap data sampel suara. Poin-poin pengujian dijabarkan sebagai berikut:

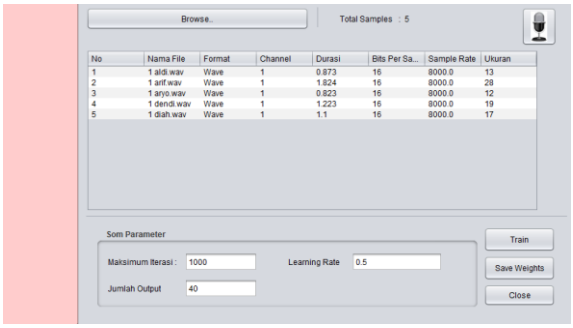
- Pengujian Fungsional
  - Pengujian untuk *use case* Ekstraksi Fitur (Penyimpanan hasil ekstraksi atau bobot)
  - Pengujian untuk *use case* Pelatihan SOM
  - Pengujian untuk use case Pengujian SOM
- Pengujian terhadap data sampel suara
  - Pengujian data suara yang sama dengan data latih.
  - Pengujian data suara yang berbeda dengan data latih.

Prosedur pengujian untuk ekstraksi fitur yaitu pertama memilih menu Pelatihan SOM, masukkan data suara lalu tekan tombol *Train*, apabila proses telah selesai maka akan keluar *alert* pemberitahuan proses selesai. Pilih tombol *Save Weights*, maka bobot hasil pembelajaran akan tersimpan.



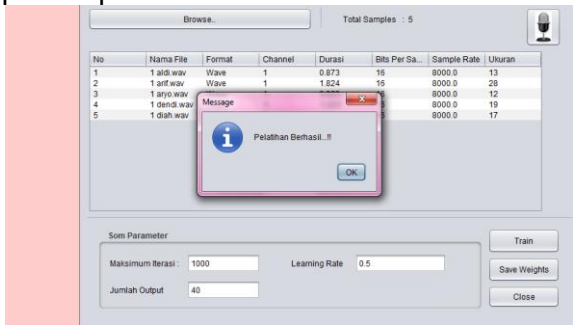
Gambar 10. Respon Penyimpanan bobot Hasil Ekstraksi

Prosedur pengujian untuk *use case* pelatihan SOM yaitu pertama masukkan data suara lalu tekan tombol *Train*.



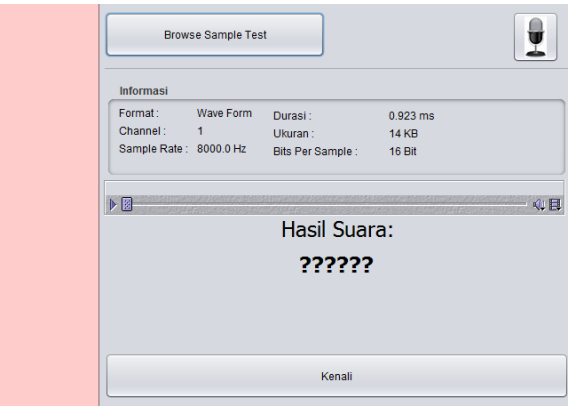
Gambar 11. Pengujian Pelatihan SOM

Apabila proses telah selesai maka akan keluar *alert* berupa pemberitahuan proses pelatihan selesai.



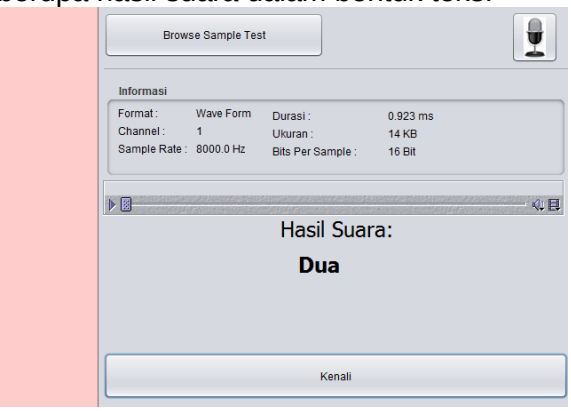
Gambar 12. Respon Pelatihan SOM

Prosedur pengujian untuk use case pengujian SOM yaitu pertama memilih menu Pengujian SOM, masukkan data suara yang akan diuji, lalu tekan tombol Kenali.



Gambar 13. Pengujian Fungsional dari Pengujian SOM

Setelah ditekan tombol Kenali kemudian sistem mengeksekusi proses pengujian yang menghasilkan keluaran berupa hasil suara dalam bentuk teks.



Gambar 14. Respon Pengujian SOM

Berdasarkan hasil pengujian fungsional yang telah dilakukan dapat disimpulkan bahwa implementasi unit dan antar muka perangkat lunak berjalan dengan baik.

Selanjutnya, pengujian terhadap data sampel suara yang merupakan lanjutan dari pengujian fungsional. Pengujian terhadap data sampel dibagi menjadi dua yaitu, pertama pengujian terhadap data suara yang sama dengan data latih. Pengujian ini menggunakan data uji yang sama dengan data latih yaitu 150 data suara untuk kata 1 sampai dengan 10. Persentase hasil pengujian untuk data yang sama dengan data latih sebesar 100% untuk data yang berhasil dikenali atau dengan kata lain keluaran dari sistem sesuai dengan apa yang diharapkan.

Tabel 1. Persentase Hasil Pengujian terhadap Data yang Berbeda dengan Data Latih

No	Data Uji untuk Kata-	Hasil Pengenalan	
		Benar	Salah
1	Satu	15	0
2	Dua	15	0
3	Tiga	15	0
4	Empat	15	0
5	Lima	15	0
6	Enam	15	0
7	Tujuh	15	0
8	Delapan	15	0
9	Sembilan	15	0
10	Sepuluh	15	0
Total		150	0
Persentase		100%	0%

Yang kedua adalah pengujian terhadap data yang berbeda dengan data latih. Pengujian ini menggunakan data uji yang berbeda dengan data latih yaitu sebanyak 50 data suara untuk kata 1 sampai dengan 10. Persentase hasil pengujian untuk data yang berbeda dengan data latih sebesar 82% untuk data yang berhasil dikenali atau keluaran dari sistem sesuai dengan apa yang diharapkan, sedangkan 18% untuk data yang tidak berhasil dikenali atau keluaran dari sistem tidak sesuai dengan apa yang diharapkan.

Tabel 2. Persentase Hasil Pengujian terhadap Data yang Berbeda dengan Data Latih

No	Data Uji untuk Kata-	Hasil Pengenalan	
		Benar	Salah
1	Satu	5	0
2	Dua	4	1
3	Tiga	2	3
4	Empat	2	3
5	Lima	3	2
6	Enam	5	0
7	Tujuh	5	0
8	Delapan	5	0
9	Sembilan	5	0
10	Sepuluh	5	0
Total		41	9
Persentase		82%	12%

Bagian ini akan menjelaskan tentang analisa dari hasil sistem yang salah mengenali suara yang diujikan. Faktor-faktor yang mempengaruhi kegagalan tersebut adalah:

1. Hasil MFCC yang lebih menyerupai pola suara lain.
2. Pengaruh data suara lain di *training set* dalam pencarian neuron pemenang.

Jaringan syaraf tiruan *Self Organizing Map* merupakan suatu algoritma *unsupervised*, maka oleh karena itu proses pengenalan tidak dapat dipengaruhi sama sekali, dan ditemukan kecendrungan suatu pola suara yang menutupi munculnya pola suara lain dalam proses pencarian *winning* neuron.

Hal tersebut menyebabkan sulitnya beberapa pola suara untuk dikenali, mengacu pada pengujian sample yang ditunjukkan pada lampiran, pola suara 2 yang diucapkan oleh Sari salah dikenali sebagai kata “dua”, dari 5 suara lain yang diucapkan oleh 5 orang yang berbeda. Untuk kata “dua” hanya bisa dikenali oleh 4 suara dari 5 suara atau memiliki akurasi 80%. Hal ini dikarenakan pola pengujian dengan kata “dua”, jaringan menghasilkan neuron pemenang yang mewakili pola kata lain.

Selain itu juga, kemiripan hasil dari fitur ekstraksi menyebabkan pengenalan menjadi salah, mengacu pada hasil pengujian sampel pada lampiran, pola suara 3 yang diucapkan oleh Yos salah dikenali sebagai kata “Tiga”. Untuk angka 3 hanya bisa dikenali 3 suara dari 5 suara yang dijadikan data uji sehingga memiliki akurasi 60%.

Pada Tabel 3. ditunjukkan nilai maksimum dari jumlah selisih nilai ekstraksi fitur dan persentasenya dari sampel-sampel yang hasilnya tidak diterima atau keluarannya tidak sesuai dengan apa yang diharapkan.

Tabel 3. Nilai maksimum Perbedaan Nilai Fitur dan Persentasenya

No	Data	Selisih Fitur	Persentase
1	Angka 2 Sari dikenali sebagai kata "Enam" dengan Angka 6 Oki sebagai kata "Enam"	0,102543012	10,25%
2	Angka 3 yang diucapkan Doni dikenali sebagai kata "Lima" dengan Angka 5 yang diucapkan Doni sebagai kata "Lima"	0,112279563	11,23%
3	Angka 3 yang diucapkan Tata dikenali sebagai kata "Enam" dengan Angka 6 yang diucapkan Ica sebagai kata "Enam"	0,110392645	11,04%
4	Angka 3 yang diucapkan Yos dikenali sebagai kata "Lima" dengan Angka 5 yang diucapkan Yos sebagai kata "Lima"	0,1193948868	11,94%
5	Angka 4 yang diucapkan Rina dikenali sebagai kata "Tiga" dengan Angka 3 yang diucapkan Sari sebagai kata "Tiga"	0,119259688	11,93%
6	Angka 4 yang diucapkan Rizki dikenali sebagai kata "Tiga" dengan Angka 3 yang diucapkan Tiara sebagai kata "Tiga"	0,076548287	7,65%
7	Angka 4 yang diucapkan Tito dikenali sebagai kata "Lima" dengan Angka 5 yang diucapkan Yos sebagai kata "Lima"	0,109663852	10,97%
8	Angka 5 yang diucapkan Rizqi dikenali sebagai kata "Sembilan" dengan Angka 9 yang diucapkan Reza sebagai kata "Sembilan"	0,080304082	8,03%
9	Angka 5 yang diucapkan Rizqi dikenali sebagai kata "Sembilan" dengan Angka 9 yang diucapkan Rani sebagai kata "Sembilan"	0,115883244	11,59%
Nilai Maksimum		0,119394886	**11,94%**

Nilai maksimum dari perbedaan nilai fitur 11,94% yang berarti kemiripan nilai fitur sebesar 88,06%. Hal ini menjadikan jaringan SOM kurang dapat melakukan klasifikasi untuk perbedaan nilai fitur MFCC

yang relatif kecil dengan batas perbedaan nilai fitur ≥ 11,94%.

KESIMPULAN

Berdasarkan hasil analisis dan implementasi yang telah dilakukan pada penelitian ini, maka dapat diambil kesimpulan sebagai berikut :

1. MFCC (*Mel Frequency Cepstral Coefficients*) dapat digunakan untuk mengembangkan jaringan SOM (*Self Organizing Maps*) untuk mengenali suara.
2. Untuk data suara yang pernah dikenali, SOM dapat dengan tepat mengenalinya.
3. Hasil akurasi yang dicapai untuk data yang sama dengan data latih sebesar 100%, sedangkan untuk data yang berbeda dengan data latih sebesar 82%.

SOM dapat dengan baik mengenali suara untuk batas perbedaan nilai fitur MFCC ≥ 11,94%.

DAFTAR PUSTAKA

[1] Aditya, R. (2012). Prototipe Pengenalan Suara Sebagai Penggerak Dinamo Starter Pada Mobil.

[2] Dewi, I. N., Firdausillah, F. & Supriyanto, C. (2013). *Sphinx-4 Indonesian Isolated Digit Speech Recognition*. *Journal of Theoretical and Applied Information Technology*. 53(1): 40-44.

[3] Dhingra, S. D., Nijhawan, G. & Pandit, P. (2013). *Isolated Speech Recognition Using MFCC and DTW*. *International Journal of Advanced Research in Electrical, Electronic & Instrumentation Engineering*. 2(8): 4085-4092.

[4] Du, X. P. & He, P. L. (2006). *The Clustering Solution of Speech Recognition Models with SOM*. *In Proceedings of the Third international conference on Advances in Neural Networks*. Heidelberg (Berlin), 150-157.

[5] Gaikwad, S. K., Gawali, B. W. & Yannawar, P. (2010). *A Review on Speech Recognition Teqhnique. International Journal of Computer Application*. 10(3): 16-24.

[6] Han, W., Chan, C. F., Choy, C. S. & Pun, K. P. (2006). *An Efficient MFCC Extraction Method in Speech Recognition*. *In Proceedings of IEEE International Symposium on Circuits & Systems*. 21-24 May. Island of Kos, 145-148.

[7] Hasan, R., Jamil, M., Rabbani, G. & Rahman, S. (2004). *Speaker Identification Using Mel Frequency Cepstral Coefficients*. 3rd International Conference on Elecrical & Computer Engineering (ICECE). 28-30 December. Dhaka (Bangladesh), 565-568.

[8] Hutchinson, M. (2004). *Windowing in Voice Conversion*. http://cnx.org/content/m12476/1.4/.

[9] Ittichaichareon, C., Suksri, S. & Yingthawornsuk, T. (2012). *Speech Recognition Using MFCC. International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012)*. 28-29 July. Pattaya (Thailand), 135-138.

[10] Ivana. (2011*).* Pengenalan Ucapan Vokal Bahasa Indonesia Dengan Jaringan Saraf Tiruan Menggunakan *Linear Predictive Coding*. Master of Science. Universitas Diponegoro, Indonesia.

[11] Kruchten, P. (2004). *The rational unified process: an introduction*. Addison-Wesley Professional.

[12] Ortega, C. A. D. L., Gonzales, M. M., Medina M. A. A., Romo, J. C. M., & Villar, V. E. G. D. (2009). *Analysis of Kohonen's Neural Network with Application to Speech Recognition. In proceeding of: MICAI 2009, Workshop Computer Vision and Pattern Recognition – WCVPR*. Juny.

[13] Pankaj, A. (2011). *Hand-Written Character Recognition Using Kohonen Network*, IJCST, 2(3), 112-115.

[14] Parmar, H. & Sharma, B. (2013). *Control System with Speech Recognition System Using MFCC and Euclidian Distance Algorithm*. *International Journal of Engineering*

Research & Technology (IJERT). 2(1): 1-5.

[15] Ping, H. (1999). *Isolated Word Speech Recognition Using Fuzzy Neural Techniques*. Doctor Philosophy, University of Windsor, Canada.

[16] Putra, D. & Resmawan, A. (2011). Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW. *Lontar Komputer*. 2(1): 8-21.

[17] Srivastava, N., Dev, H. & Abbas, Q. (2013). *Speech recognition using MFCC and Neural Network*. *National Conference on Challenges & Opportunities for Technological Innovation AIMT*. February. India.

[18] Swamy, S., Radhika, M. V., Roopa, G. & Ramakrishnan, K. V. (2013). Speaker Independent Digit Recognition System. In Proceeding of 47th Annual National Convention Of Computer Society Of India Organized. February. India.

[19] Taner, M.T. (1997). *Kohonen's Self Organizing Networks with "Conscience". In Rock Solid Images complete document with pp*. November. 1-7.

[20] Venkateswarlu, R. L. K., Raviteja, R. & Rajeev, R. (2012). *The Performance Evaluation of Speech Recognition by Comparative Approach, Advances in Data Mining Knowledge Discovery and Applications*. http://www.intechopen.com/books/advances-in-data-mining-knowledge-discovery-and-applications/the-performance-evaluation-of-speech-recognition-by-comparative-approach.