ANALISIS DAN IMPLEMENTASI APLIKASI PENGENALAN SUARA MENJADI TEKS MENGGUNAKAN METODE JARINGAN SYARAF TIRUAN BACKPROPAGATION

ANALYSIS AND IMPLEMENTATION OF SPEECH TO TEXT APPLICATION USING BACKPROPAGATION NEURAL NETWORK

Rayani Budi Andhini[1] , Budhi Irawan,S.Si. ,MT.[2] , Inung Wijayanto,ST.,MT.[3]

[1,2,3]Fakultas Teknik Elektro, Universitas Telkom

[1]rayanib.andhini@gmail.com, [2]bir@ittelkom.ac.id, [3]iwijayanto@telkomuniversity.ac.id

Abstrak

Bahasa merupakan interaksi yang paling natural yang digunakan manusia. Manusia dengan mudah menggunakan Bahasa sebagai alat untuk berkomunikasi satu sama lain. *Speech recognition* merupakan suatu metode yang dikembangkan untuk komunikasi manusia dengan mesin menggunakan suara. Pada sistem pengenalan suara yang dibangun dalam penelitian ini, metode ekstraksi ciri yang digunakan adalah *Discrete Cosine Transform*. Untuk proses klasifikasi, yang digunakan dalam pembuatan sistem ini adalah metode Jaringan Syaraf Tiruan Backpropagation. Sistem ini mengidentifikasi 30 kata dengan total data latih 270 dan untuk pengujian dengan total data uji 180. Berdasarkan hasil pengujian dengan parameter jumlah hidden neuron 200, jumlah nilai ciri 500 dan learning rate 0.02, yang diujikan *non realtime* diperoleh bahwa sistem pengenalan suara menjadi teks sebesar 51%.

Kata kunci : s*peech recognition, Discrete Cosine Transform,* Jaringan Syaraf Tiruan *Backpropagation*

Abstract

Language is the most natural interaction human use. Humans easily use language as a tool to communicate with each other. Speech recognition is a method developed for human to communicate with machine using sound. In the speech recognition system for this research, Discrete Cosine Transform is used for feature extraction method. For classification, the method used in this research is Backpropagation Neural Network. This system identified 30 classes with a total 270 training data and with a total of 180 for testing data. Based on the results, with the number of hidden neuron 200, the number of feature value 500 and learning rate 0.02, for data tested in non realtime this research resulted in 51% accuracy.

Key Word :* speech recognition, Discrete Cosine Transform, Backpropagation Artificial Neural Networ*k

1. Pendahuluan

Teknologi *speech recognition* dapat dikembangkan sehingga suara yang diucapkan dapat diterjemahkan menjadi kata pada komputer. Teknologi penerjemah suara ke komputer atau *speech recognition*, akan berperan sangat penting pada perkembangan teknologi yang akan datang. Akan banyak aplikasi yang dikembangkan berbasis *speech recognition*[7], sehingga komunikasi masukan ke komputer tidak lagi hanya menggunakan perangkat keras, seperti *keyboard* dan *mouse*.

Maka pada penelitian ini, akan dibuat suatu aplikasi yang dapat menerjemahkan suara ke dalam bentuk teks. Penelitian ini terdiri dari tiga tahapan utama yaitu preproses sinyal suara masukan, ekstraksi ciri dan klasifikasi. Metode ekstraksi ciri yang akan digunakan disini adalah *Discrete Cosine Transform* (DCT). Kemudian untuk klasifikasi akan digunakan metode Jaringan Syaraf Tiruan Backpropagation. Metode Jaringan Syaraf Tiruan Backpropagation akan mengklasifikasikan setiap suara dengan sebuah kata dan akan disimpan di *database*. Pada metode pencarian suara yang masuk akan dibandingkan dengan yang sudah ada di *database*, kemudian keluaran yang didapatkan akan ditampilkan dengan bentuk teks.

2. Dasar Teori

2.1 Discrete Cosine Transform

Discrete Cosine Transform merupakan suatu teknik untuk merubah suatu sinyal menjadi komponen frekuensi dasar, dengan memperhitungkan nilai riil dari hasil transformasi. *Discrete Cosine Transform* merupakan transformasi yang berhubungan dengan transformasi Fourier yang memberikan fungsi diskrit dengan hanya mengambil nilai *cosinus* dari eksponensial kompleks.

Discrete Cosine Transform adalah transformasi ideal untuk proses kompresi data, dengan kemampuan memampatkan data hingga 99% lebih kecil dari sinyal masukannya[5]. DCT dapat digunakan untuk mendapatkan nilai koefisien. Hal ini dikarenakan DCT mampu menghasilkan sinyal berfrekuensi rendah lebih sedikit dan frekuensi tinggi yang banyak. DCT merekonstruksi urutan data dengan DCT koefisien, parameter yang berguna yang digunakan untuk reduksi data[4]. Persamaan DCT dapat dilihat pada persaman di bawah.

$$\text{[illegible]} = \text{[illegible]} \sum_{\text{[illegible]}=1}^{N} \text{[illegible]} \cos \frac{\text{[illegible]}(2\text{[illegible]}-1)(\text{[illegible]}-1)}{2N}, \quad k = 1, \ldots, N \tag{2.1}$$

Dimana

$$\text{[illegible]} = \begin{cases} \frac{1}{\sqrt{N}} & k = 1 \\ \sqrt{\frac{2}{N}} & 2 \le k \le N \end{cases}$$

2.2 Jaringan Syaraf Tiruan Backpropagation

Jaringan syaraf tiruan (JST) merupakan sistem adaptif yang dapat mengubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut[6]. JST juga dapat digunakan untuk mengelola hubungan input dan output untuk menemukan pola-pola data. Ada dua metode untuk menemukan pola-pola data ini, yaitu *unsupervised learning* dan *supervised learning*[1][8].

JST Backpropagation merupakan metode *supervised learning*, dimana sudah ditentukan pola output yang diinginkan. Proses latih dalam metode JST Backpropagation melibatkan tiga tahap, yaitu tahap *feedforward* untuk melatih *pattern* data input, kalkulasi dan backpropagation dari error, dan tahap penyesuaian bobot[1][3]. Arsitektur yang digunakan dalam JST Backpropagation bisa berupa *single-layer net* atau *multilayer net*. Arsitektur yang digunakan adalah *multilayer net* dengan satu *hidden layer*.

Pada proses *feedforward* setiap unit input (X_i) mendapatkan nilai sinyal input dan melanjutkan sinyal ke *hidden unit* (Z_j). Setiap *hidden unit* akan melakukan menghitung nilai aktivasi z_j dan mengirim sinyal hasil ke setiap unit output. Setiap output unit (Y_k) akan menghitung nilai aktivasi y_k untuk membentuk hasil sesuai dengan output target. Selama proses latih setiap hasil putput akan disbandingkan dengan output target untuk menentukan nilai error. Berdasarkan nilai error, bobot dari setiap sinapsis akan diperbarui, dilakukan secara mundur. Proses tersebut dilakukan hingga nilai *error* sudah memenuhi syarat yang telah ditentukan atau syarat *epoch* sudah terpenuhi. Parameter yang terdapat pada JST adalah sebagai berikut:

a. x_i = nilai input
b. u_{ij} = bobot sinapsis antara x_i dan z_j
c. z_i = nilai *hidden*
d. w_{ij} = bobot sinapsis antara z_i dan y_j
e. v_{yi} = hasil *summing function* di y_i
f. v_{zi} = hasil *summing function* di z_i
g. ey_i = *error* di y_i
h. ez_i = *error* di z_i
i. dy_i = *error correction* di y_i
j. dz_i = *error correction* di z_i
k. α = *learning rate*

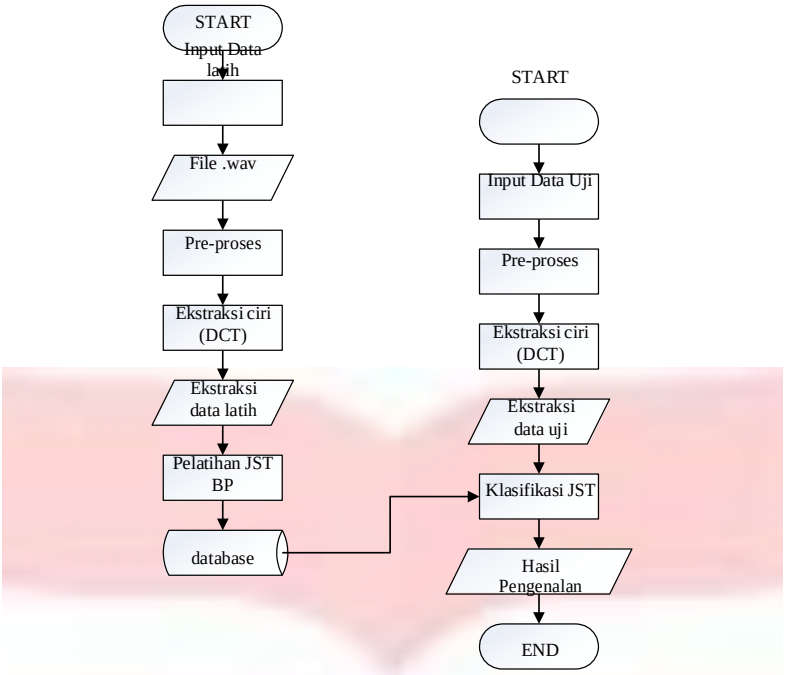
3. Perancangan Sistem

Diagram alir sistem dapat digambarkan seperti pada gambar 3.1. Ada tiga proses utama yang dilakukan, yaitu preproses, ekstraksi ciri dengan metode *Discrete Cosine Transform* dan klasifikasi menggunakan Jaringan Syaraf Tiruan (JST) Backpropagation. Proses klasifikasi pada JST berupa proses *training* dan proses *testing*.

Sebelum langkah preproses, sura direkam terlebih dahulu. Akuisisi data suara pada penelitian ini direkam dengan menggunakan *Software* 'Audacity'. Suara direkam dengan *frequency sampling* 8000Hz[2][3], dalam bentuk wav format.. Sinyal input suara yang telah direkam akan diproses agar mendapatkan bentuk sinyal yang diinginkan untuk proses selanjutnya, yaitu dengan cara normalisasi sinyal suara sehingga memiliki nilai maksimum dan minimum pada *range* 1 dan -1, dan *cropping*, yaitu memotong jeda kosong di awal dan akhir sehingga informasi yang dimiliki lebih efisien.

Discrete Cosine Transform digunakan sebagai metode ekstraksi ciri. Ekstraksi ciri dilakukan untuk mendapatkan informasi penting dari sebuah sinyal suara sehingga dapat dibedakan suara satu dengan suara yang

lainnya. Proses ekstraksi ciri dilakukan pada setiap sinyal suara yang memiliki nilai koefisien berbeda. Koefisien ciri yang telah didapatkan akan disimpan dalam format .csv.

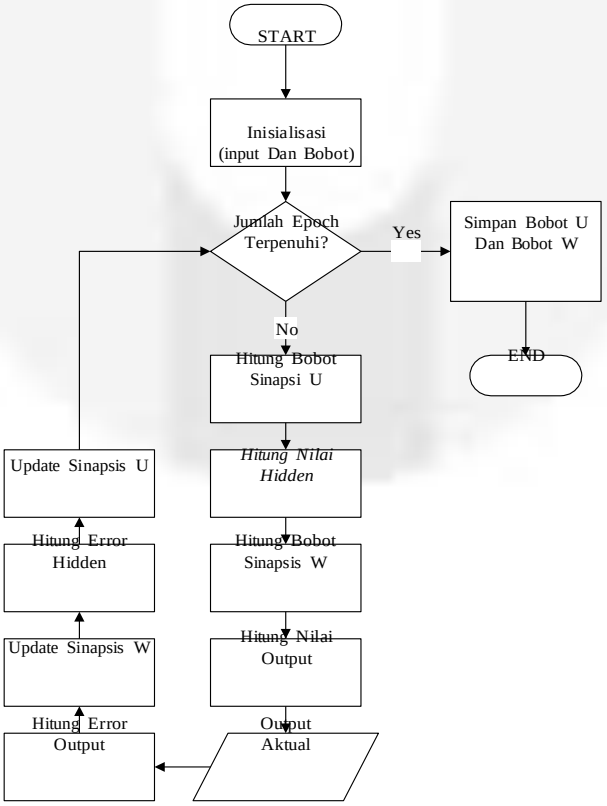


Gambar 3. 1 diagram alir

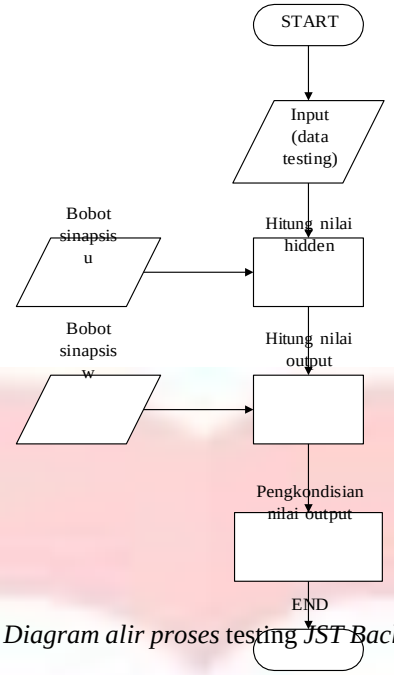
3.1 Backpropagation

Proses latih JST Backpropagation adalah untuk mencari bobot tiap sinapsis dengan nilai *error* terkecil. Untuk proses latih JST Backpropagation dilakukan proses *feedforward* dan *backward*. Diagram alir pada gambar 3.2 menjelaskan bagaimana proses *training* JST Backpropagation.

Untuk proses *testing*, proses JST dilakukan dengan menggunakan bobot yang sudah dilatih sebelumnya. Proses yang dilakukan hanyalah proses *feedforward*. Proses *testing* dapat dijelaskan dengan diagram alir di gambar 3.3.



Gambar 3. 2 Diagram alir training JST Backpropagation



Gambar 3. 3 Diagram alir proses testing *JST Backpropagation*

3.2 Identifikasi Data

Proses identifikasi data dilakukan secara *real time* dan *non real time*. Pada proses uji, tahap preproses dan ekstraksi ciri tidak berbeda dengan proses latih. Untuk tahap klasifikasi JST Backpropagation, parameter-parameter yang digunakan adalah nilai yang didapatkan dari hasil proses latih. Perbedaannya proses JST backpropagation yang dilakukan hanyalah proses *feedforward* saja.

Untuk proses identifikasi data *non real time*, data atau suara sebelumnya sudah direkam terlebih dahulu. Identifikasi data *non real time* dilakukan untuk pengecekan nilai bobot dan akurasi dari bobot tersebut bila dilakukan proses uji dengan data uji yang telah disediakan. Ada 30 kelas data latih yang akan disimpan dalam *database* aplikasi ini (Tabel 3.1).

Tabel 3. 1 Klasifikasi data

No	Kelas	No	Kelas	No	Kelas
1	Besar	11	Jahat	21	Paham
2	Bola	12	Karena	22	Pakai
3	Badminton	13	Kami	23	Sangat
4	Cantik	14	Kamu	24	Sedih
5	Dengan	15	Keranjang	25	Sandang
6	Halaman	16	Kabut	26	Terimakasih
7	Ilmu	17	Kabar	27	Tampak
8	Inap	18	Makan	28	Ucap
9	Ingat	19	Nikah	29	Pelangi
10	ikut	20	ombak	30	Takut

4. Pengujian dan Analisis Parameter Jaringan

Pada bab ini akan dilakukan pengujian dan analisis terhadap program dengan beberapa skenario yang telah dirancang.

4.1 Pengujian dan Analisis Pengaruh *Learning Rate*

Pada pengujian ini, nilai parameter *learning rate* yang akan diujikan adalah 0.01; 0.02; dan 0.03. *Hidden neuron* yang digunakan 50, dengan maksimum epoch 10000 dan batas toleransi error 10^{-4}. *Threshold* yang digunakan pada jaringan adalah 0.2. Total data yang akan diujikan berjumlah 180 sampel, dari 30 kata. Hasil yang didapatkan dijelaskan pada tabel di bawah ini.

Tabel 4. 1 Pengaruh Learning Rate terhadap Akurasi

Learning Rate	Akurasi (%)	
	Data Latih	Data Uji
0.01	38	20
0.02	39	22
0.03	27	11

Pada tabel 4.1 dapat disimpulkan bahwa *learning rate* terbaik yang didapatkan adalah 0.02. Nilai *learning rate* mempengaruhi tingkat akurasi dari sistem. Semakin kecil nilai *learning rate*, akan semakin teliti sistem dalam mencari error. Akan tetapi, pemilihan nilai *learning rate* yang terlalu besar atau terlalu kecil akan membuat sistem menjadi tidak stabil.

4.2 Pengujian dan Analisis Pengaruh *Hidden Neuron*

Pada pengujian ini diuji pengaruh jumlah *hidden neuron* pada JST. Pengujian ini dilakukan dengan merubah neuron pada jaringan. Neuron yang akan diuji adalah 500, 100, dan 200. *Threshold* yang dipakai 0.2. parameter lainnya yaitu, *learning rate* 0.02, maksimum epoch 10000 dan batas toleransi error 10^{-4}. Pada tabel di bawah dapat dilihat hasil pengujian tiap neuron.

Tabel 4. 2 Pengaruh Hidden Neuron Terhadap Akurasi

Hidden Neuron	Akurasi (%)		MAE	Epoch ke
	Data Latih	Data Uji		
50	38	16	0.1646	10000
100	98	46	$2.14x10^{-5}$	9030
200	99	51	0.0001	6484

Penambahan jumlah *hidden neuron* membuat pelatihan jaringan syaraf tiruan backpropagation menjadi lebih mudah. Dari pengujian dapat disimpulkan semakin banyak jumlah hidden neuron, hasil dari pelatihan akan semakin teliti dikarenakan kalkulasi yang dilakukan di layer hidden menjadi bertambah. Pada tabel 4.2 dapat disimpulkan *hidden neuron* 200 memiliki akurasi tertinggi yaitu sekitar 51%.

4.3 Pengujian dan Analisis Pengaruh Jumlah *Feature*

Pada pengujian ini diuji pengaruh jumlah *feature* (ciri) terhadap akurasi. Jumlah ciri yang akan diuji yaitu 500, 700, dan 1000. Parameter lainnya yaitu *learning rate 0.02*, *hidden neuron* 200, *threshold* 0.2. Maksimum epoch pelatihan 10000 dan batas tolerasndi error 10^{-4}. Hasil pengujian dapat dilihat pada tabel di bawah.

Tabel 4. 3 Pengaruh Jumlah Ciri Terhadap Akurasi

Jumlah nilai ciri	Akurasi (%)		MAE	Epoch ke
	Data Latih	Data Uji		
500	99	51	0.0001	6848
700	98	15	0.00009	7113
1000	99	10	0.0005	10000

Pada tabel 4.3 dapat disimpulkan bahwa jumlah nilai ciri yang paling cocok adalah 500 yang menghasilkan akurasi terbaik yaitu 51%. Jumlah nilai ciri yang menghasilkan akurasi yang paling buruk adalah 1000. Dapat dilihat pada tabel 4.3, pada maksimum epoch 10000 total error yang didapat 0.0005, belum memenuhi syarat dari batas toleransi error. , karena semakin besar nilai ciri yang digunakan sebagai input pada pelatihan jaringan syaraf tiruan backpropagation semakin besar proses yang harus dilakukan. Maka dari itu maksimum epoch 10000 tidak cukup untuk dilakukan pelatihan. Agar didapatkan hasil yang lebih ideal, maksimum epoch harus diperbesar atau ditambahkan *hidden neuron.*

4.4 Pengujian Waktu Komputasi

Pengujian ini dilakukan untuk mengukur waktu komputasi sistem pengenalan suara menjadi teks tiap kelas dengan menggunakan parameter terbaik dari pengujian sebelumnya. Parameter yang digunakan yaitu, *learning rate* 0.02, *hidden neuron* 200, jumlah nilai ciri 500. Hasil pengujian dapat dilihat pada tabel di bawah ini.

Tabel 4. 4 Rata-Rata Waktu Komputasi

Kelas	Waktu komputasi rata-rata	Kelas	Waktu komputasi rata-rata	Kelas	Waktu komputasi rata-rata
Besar	3921.3	Kabar	2099.7	Jahat	2173.3
Bola	1859.7	Makan	2629.5	Karena	2475.7
Badminton	6740.0	Nikah	2250.8	Kami	1991.8
Cantik	1927.8	ombak	2031.7	Kamu	2149.2
Dengan	2777.8	Paham	2418.5	Keranjang	3690.3
Halaman	4404.5	Pakai	2418.5	Tampak	1632.3
Ilmu	2027.5	Sangat	3170.8	Ucap	2323.0
Inap	1702.2	Sedih	1849.5	Pelangi	3868.0
Ingat	1410.8	Sandang	2689.0	Takut	2214.5
Ikut	1929.7	Terimakasih	5191.3	Kabut	2150.3

Tabel 4.4 menampilkan rata-rata waktu komputasi dari tiap kata. Waktu terkecil dari tiap komputasi yaitu 1410.8 ms dan terbesar6740 ms. Untuk kecepatan komputasi rata-rata keseluruhan kata yaitu 2700.47 ms.

5. Kesimpulan dan Saran

Dari hasil pengujian dan analisa yang telah dilakukan pada sistem, dapat diambil beberapa kesimpulan sebagai berikut :

1. Dalam perancangan system pengenalan suara menjadi teks ada beberapa hal yang perlu diperhatikan yaitu:

a. Nilai *learning rate* mempengaruhi tingkat akurasi dan lama waktu komputasi dari jaringan. Pemilihan nilai *learning rate* yang tepat akan meningkatkan nilai akurasi.
b. Penambahan jumlah *hidden neuron* dapat meningkatkan akurasi dari jaringan, tetapi akan menambah waktu komputasi dari pelatihan.
c. Semakin besar jumlah nilai ciri maka akan menambah dimensi input dari JST, maka dari itu akan proses pelatihan akan lebih lama. Sehingga, apabila dimensi input yang terlalu besar dapat menyebabkan tingkat akurasi menurun.

2. Sistem pengenalan suara dapat mengenali kata dengan maksimal akurasi 51% dengan 270 data latih dan 180 data uji.
3. Spesifikasi parameter yang memberikan hasil yang paling optimal adalah jumlah *hidden neuron* 500, *learning rate* 0.02, panjang dari nilai ciri 500.

Untuk pengembangan lebih lanjut penulis memberikan beberapa saran, antara lain:

1. Dapat dibuat sistem dengan tingkat akurasi yang lebih baik dengan perbaikan suara masukan pada saat *preprocessing*.
2. Perbaikan ekstraksi ciri untuk suara masukan dengan menggunakan metode yang lebih baik.
3. Sistem dapat dijalankan secara *realtime.*

Daftar Pustaka

[1] Fausett, Lauren V. 1993. Fundamental of Neural Network: Architectures, Algorithm And Application. Pearson.
[2] Joshi, Siddhant C. 2014. MATLAB Based Back-Propagation Neural Network for Automatic Speech Recognition. IJAREEIE. Volume 3 - No.7.
[3] Putra, Andika B.. 2007. Speech Recognition Menggunakan Gabor Wavelet Dan Jaringan Saraf Tiruan Backpropagation Untuk Sistem Keamanan Berbasis Suara. SNSI Bali.
[4] R. B. Shinde, Dr. V. P. Pawar. 2012. Vowel Classification based on LPC and ANN. IJCA, Volume 50 – No.6.
[5] Setyawan, M. Taufik. 2011. Simulasi Tapis Finite Impulse Response (Fir) Dengan Discrete Cosine Transform (DCT). Universitas Diponegoro.
[6] Siang, J.J.. 2005. Jaringan Saraf Tiruan dan Pemrogramannya Menggunakan Matlab. ANDI Yogyakarta, Yogyakarta.
[7] Speaker Independent Connected Speech Recognition. [Online]. Available: http://www.fifthgen.com/speaker-independent-connected-s-r.htm. [Diakses 3 Februari 2015]
[8] Srivastava, Nidhi. 2014. Speech Recognition using Artificial Neural Network. IJESIT. Volume 3 - No.3.