

## PENGENALAN SUARA MANUSIA MENGGUNAKAN METODE NPC DENGAN JST-BP

Nindiyanto Adhi Chandra<sup>1</sup>, Iwan Iwut Tritoasmoro<sup>2</sup>, Achmad Rizal<sup>3</sup>

<sup>1</sup>Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

---

### Abstrak

Informasi sinyal suara mempunyai karakteristik yang unik. Berbagai penelitian dalam pengolahan sinyal suara pun telah banyak dikembangkan. Salah satu hasil penelitian tersebut adalah aplikasi speaker identification. Speaker Identification adalah suatu proses mengenali pembicara (client) secara otomatis melalui sinyal suara dari pembicara tersebut sebagai informasinya. Biasanya sistem speaker identification menggunakan fonem sebagai unit pengenalannya. Namun masih terdapat kesulitan dalam mengenali kelas-kelas fonem tertentu yang mempunyai durasi yang cukup singkat.

Berdasarkan permasalahan di atas maka dikembangkan suatu teknik pengenalan suara berbasis software menggunakan database diphone sebagai unit pengenalannya. Secara akustik, diphone lebih mudah dikenali karena memberikan informasi akustik yang lebih banyak daripada fonem. Hasil dari penelitian Tugas Akhir ini menunjukkan bahwa ekstraksi feature suara jantung menggunakan LPC memberikan hasil yang cukup baik karena mampu menunjukkan kemiripan feature dari sinyal suara yang digunakan sebagai database referensi. Berdasarkan hasil pengujian dengan menggunakan 90 data pengujian metoda JST-BP memiliki akurasi pengujian 80% dengan blindset database. Berdasarkan hasil penelitian tersebut didapatkan konfigurasi optimal dari jaringan saraf tiruan adalah 250 hidden neuron, learning rate 0.09 dan momentum 0.35

Kata Kunci : speaker identification, Linier Predictive Coding (LPC), jaringan saraf

---

### Abstract

Speech signal information has a unique characteristic. Many researches in the speech signal processing field has done. One of the results produced is speaker identification application. Speaker Identification is a process of determining the speaker provides a given utterance. Generally, phonemes used in the speaker identification system as a recognizer unit, but the problem is located in the certain phonemes classes which has a short duration. In order to solve this problem, speech recognition method using diphone (phonemes combination) database as an alternate recognition unit is developed.

In this final project, diphones used as a reference database in the speech recognition system is being explored. The diphone itself is easier to be recognized than the phoneme, since it gives more acoustics information. The identification method used here is Back Propagation Neural Network. The method used in speech signal feature extraction is Linear Predictive Coding (LPC). Result in this final assignment showed that the feature extraction method using LPC give a good result because it show the resemblances of the features from the speech signal which is used for reference database. Based on the test, Back propagation method has an accuracy 80% using blindest database. Based on the research result we obtained the optimum configuration from neural network are 250 hidden neurons, learning rate 0.09, and momentum 0.35

Keywords : speaker identification, Linier Predictive Coding (LPC), jaringan saraf

---

---

# BAB I

## PENDAHULUAN

### 1.1.Latar Belakang Masalah

Informasi sinyal suara mempunyai karakteristik yang unik. Berbagai penelitian dalam pengolahan sinyal suara pun telah banyak dikembangkan. Salah satu hasil penelitian tersebut adalah aplikasi *speaker recognition*. *Speaker Recognition* adalah suatu proses mengenali pembicara (*client*) secara otomatis melalui sinyal suara dari pembicara tersebut sebagai informasinya. Biasanya sistem *speaker recognition* menggunakan fonem sebagai unit pengenalannya. Namun masih terdapat kesulitan dalam mengenali kelas-kelas fonem tertentu yang mempunyai durasi yang cukup singkat.

Berdasarkan permasalahan di atas maka dikembangkan suatu teknik pengenalan suara berbasis *software* menggunakan database *diphone* sebagai unit pengenalannya. Secara akustik, *diphone* lebih mudah dikenali karena memberikan informasi akustik yang lebih banyak daripada fonem. Masukan *software* berupa kata dalam bahasa Indonesia yang telah direkam sebelumnya.

Untuk mendapatkan informasi dari sinyal suara perlu dilakukan ekstraksi ciri sehingga dapat dianalisis untuk tiap variasi sinyal suara yang ada. Dari ciri yang ada pada setiap *diphone* dicoba untuk mengenalinya dan melakukan identifikasi terhadap pembicara. Dalam proses pengenalan diperlukan suatu metode pemodelan dan sistem optimasi yang dapat mendukung proses pengenalan sehingga tidak terjadi kesalahan pengertian yang disebabkan dalam penggunaan metode pengenalan. Karena model sistem vokal manusia adalah nonlinear, maka pada tugas akhir ini dibuat suatu sistem identifikasi suara manusia dengan menggunakan nonlinear predictor, yaitu dengan menggunakan metode *Neural Predictive Coding* (NPC) dan *Jaringan Saraf Tiruan Propagasi Balik* (JST-BP). Rasio yang diharapkan adalah 80%-90%.

---

*Pengenalan Suara Manusia Menggunakan Metode NPC Dengan JST-BP*

## 1.2. Tujuan

1. Mensimulasikan dan menganalisa aplikasi *speaker identification* pada sistem *speech recognition*.
2. Membuktikan apakah metode *Neural Predictive Coding* adalah algoritma *feature extraction* yang baik dengan indikasi rasio pengenalan sinyal ciri yang baik pula.
3. Menganalisa kinerja sistem *speaker identification* melalui rasio pengenalan sinyal ciri dari algoritma *Neural Predictive Coding* dan sistem deteksi pola ciri melalui *Jaringan Saraf Tiruan Propagasi Balik (JST-BP)*.

## 1.2. Rumusan Masalah

Rasio pengenalan sistem *speaker identification* secara teknis salah satunya ditentukan oleh algoritma pengkodeannya. Karena pemodelan sistem produksi ucapan manusia bersifat nonlinear, maka untuk meningkatkan rasio pengenalan digunakan *nonlinear extraction*. *Neural Predictive Coding* adalah salah satu algoritma pengkodean nonlinear. Tugas akhir ini membuktikan apakah *Neural Predictive Coding* memiliki rasio pengenalan yang baik atau tidak? Dan apakah simulasi sistem *speaker identifikasi* berjalan dengan baik atau tidak?

## 1.3. Batasan Masalah

Batasan masalah pada tugas akhir ini adalah :

1. Masukan perangkat lunak yang dirancang berupa kata dasar dalam bahasa Indonesia yang telah direkam dan disimpan dalam format file .WAV pada frekuensi *sampling* 8 kHz, PCM mono, resolusi 16 bit.
2. Unit pengenalan yang digunakan adalah *diphone* dengan kombinasi konsonan-vokal (KV). Database *diphone* diambil dari sampel suara tiga orang laki-laki saja, dimana tiap pembicara mengucapkan 30 kata.
3. Database *diphone* yang digunakan sudah pernah digunakan sebelumnya untuk topik tugas akhir yang berbeda.
4. Metode untuk menentukan batas-batas *diphone* dalam suatu kata menggunakan selubung energi dan *zero-crossing rate*.
5. Ekstraksi ciri menggunakan metode *Linear Predictive Coding (LPC)*

---

*Pengenalan Suara Manusia Menggunakan Metode NPC Dengan JST-BP*

6. Deteksi menggunakan Jaringan Saraf Tiruan Propagasi Balik yang selanjutnya disebut JST-BP saja.
7. Pada tugas akhir ini dilakukan analisa terhadap kinerja jaringan saraf tiruan terhadap sistem pengenalan suara yang dibuat
8. Bahasa pemrograman yang digunakan adalah MATLAB 7.1

#### 1.4. Metodologi Penelitian

Metodologi yang digunakan dalam penyusunan tugas akhir ini adalah :

1. Studi literatur

Studi literatur dilakukan untuk memahami konsep-konsep dasar yang berkaitan dengan :

- Sistem *speech recognition*
- Metode *Neural Predictive Coding (NPC)*
- Jaringan Saraf Tiruan Propagasi Balik

Studi literatur dilakukan melalui jurnal ilmiah, buku, dan *internet*.

2. Konsultasi dan Bimbingan

Konsultasi dilakukan dengan dosen pembimbing dan juga diskusi dengan beberapa teman dan pihak lain yang kompeten

3. Pengumpulan sampel

Pengumpulan sampel dilakukan melalui proses perekaman suara dari 3 orang pembicara, dimana tiap pembicara mengucapkan 30 kata. Hasil rekaman ini kemudian dipotong-potong untuk memperoleh sinyal *training* dan sinyal uji.

4. Implementasi sistem

Sistem diimplementasikan melalui perangkat lunak menggunakan bahasa pemrograman MATLAB 7.1.

5. Analisa data

Data-data hasil simulasi dianalisa untuk memperoleh performansi sistem *speech recognition* dalam mengenali suara, serta menganalisa performansi metode Neural Predictive Coding (NPC) dan Jaringan Saraf Tiruan Propagasi Balik.

### 1.5. Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini adalah sebagai berikut:

#### **BAB I : Pendahuluan**

Bab ini menguraikan latar belakang pemilihan topik tugas akhir ini, masalah yang dihadapi, batasan-batasan yang ditetapkan berkaitan dengan masalah yang ada, dan tujuan tugas akhir ini.

#### **BAB II : Dasar Teori**

Bab ini berisi dasar teori dari metoda-metoda yang digunakan dalam tugas akhir ini, diantaranya adalah sistem produksi suara manusia, *Neural Predictive Coding (NPC)*, *Linear Predictive Coding*, dan *Jaringan Saraf Tiruan Propagasi Balik (JST-BP)*.

#### **BAB III : Perancangan Sistem**

Bab ini menggambarkan rancangan dari sistem dan menguraikan penjelasan dari tiap blok dari rancangan.

#### **BAB IV : Simulasi dan Analisa Sistem**

Pada bab ini memberikan gambaran hasil simulasi dan analisa dari hasil simulasi yang telah dilakukan.

#### **BAB V : Kesimpulan dan Saran**

Bab ini memberikan kesimpulan mengenai hal-hal yang telah dibahas dalam tugas akhir ini dan saran-saran untuk pengembangan topik tugas akhir ini.



Telkom  
University

---

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan hasil analisa sistem pada bab sebelumnya, maka dapat diambil kesimpulan sebagai berikut :

1. Ekstraksi *feature* sinyal suara menggunakan LPC memberikan hasil yang cukup baik karena mampu menunjukkan kemiripan *feature* dari sinyal suara
2. Pelatihan jaringan dengan mengubah-ubah jumlah hidden neuron pada hidden layer dapat mempengaruhi kinerja jaringan. Dari percobaan didapat, semakin besar jumlah hidden neuron maka semakin sedikit iterasi yang dilakukan untuk mencapai target error yang diinginkan tetapi waktu untuk mencapai target error menjadi lebih lama. Jumlah hidden neuron yang optimal adalah 250 hidden neuron dengan jumlah epoch 8055 dan waktu untuk mencapai MSE yang diinginkan adalah 307.0341 detik.
3. Pelatihan jaringan dengan mengubah-ubah nilai *learning rate* dapat mempengaruhi kinerja jaringan. Dari percobaan didapat, semakin besar nilai *learning rate* maka semakin sedikit iterasi yang dilakukan untuk mencapai target error yang diinginkan. Nilai *learning rate* yang optimal dari percobaan didapat 0.09 dengan jumlah epoch 383 dan waktu untuk mencapai MSE yang diinginkan adalah 15.5324 detik.
4. Semakin besar momentum maka akan semakin sedikit iterasi yang dilakukan karena dengan bertambahnya momentum maka perubahan data yang mencolok dapat dihindari sehingga perubahan bobot dapat dilakukan secara cepat. Namun perlu diperhatikan juga penambahan momentum yang terlalu besar dapat mempengaruhi kinerja jaringan karena dapat memperbesar kemungkinan terjadinya error. Dari percobaan didapat nilai optimal momentum adalah 0.35
5. Berdasarkan hasil pengujian dapat direkomendasikan struktur jaringan saraf tiruan yang optimal adalah dengan 250 hidden neuron, *learning rate* 0.09, dan momentum 0.35
6. Berdasarkan hasil pengujian, metoda JST-BP memiliki akurasi pengujian 80% dengan melakukan pengenalan terhadap *blind-set* database.

---

*Pengenalan Suara Manusia Menggunakan Metode NPC Dengan JST-BP*

## 5.2. Saran

1. Perlunya digunakan metode *feature extraction* sinyal yang lain untuk perbaikan dalam mengekstraksi parameter-parameter sinyal sehingga ekstraksi parameter-parameter suara dapat menghasilkan *feature* yang lebih baik. Selain itu dapat dijadikan perbandingan dengan metode yang telah penulis gunakan
2. Perlu diperbanyak data yang digunakan sebagai data latih dan data uji. Termasuk menggunakan database referensi dengan gender yang berbeda.
3. Menggunakan metode pengenalan JST yang lainnya dalam mengenali pola suara.
4. Adanya kemungkinan untuk mengimplementasikan secara *real time* untuk meningkatkan akurasi dari sistem



## DAFTAR PUSTAKA

- [1] Supriyatno, 2007, *Speech-to-text Menggunakan Independent Component Analysis (ICA), Pitch Conversion, dan Hidden Markov Model (HMM)*, Tugas Akhir, Teknik Elektro, STT Telkom: Bandung.
- [2] Lincoln, Mike. 1999. *Characterization of Speakers for Improved Automatic Speech Recognition*. School of Information Systems. Univeristy of East Anglia, Norwich.
- [3] Bastien, Patrick. Pitch shifting and voice transformation techniques. TC Helicon Vocal Technologies.
- [4] Saldana, Isaac and Ginsberg, David. *Remote Speaker and Speaker Recognition A Senior Design Project*. Department of Electrical Engineering University of California, Riverside.
- [5] Kusumadewi, Sri. 2004. *Membangun Jaringan Saraf Tiruan menggunakan Matlab dan EXCEL LINK*. Yogyakarta : Graha Ilmu.
- [6] M. Chetouani, M. Faundez-Zanuy, B. Gas, J.L. Zarader. *A New Nonlinear Feature Extraction Algorithm For Speaker Verification*. Universite De Paris VI, Paris, France.
- [7] Jarwadi, 2007, *Speech-to-text Menggunakan Database Diphone Dalam Bahasa Indonesia Dengan Metode Pendekatan Hybrid Hidden Markov Model Dan Algoritma Genetika*, Tugas Akhir, Teknik Elektro, STT Telkom: Bandung.

Telkom  
University