Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)
Edisi. .. Volume. .., Bulan 20.. ISSN : 2089-9033

Pengidentifikasian Suara Tinggi Dan Rendah Pada Pria Dan Wanita Ditinjau Dari *Pitch* Dan *Formant* Menggunakan Metode *Support Vector Machine*

Indra Tri Prabowo

Teknik Informatika - Universitas Komputer Indonesia
Jl. Dipati Ukur No. 112-114 Bandung
Email : indra3p@gmail.com

ABSTRAK

Pengenalan suara merupakan proses identifikasi suara berdasarkan kata yang diucapkan seseorang yang ditangkap oleh perangkat *input* suara untuk dikenali dan kemudian diterjemahkan menjadi sebuah data yang dipahami oleh komputer. Ketika manusia mengeluarkan suara, saat itulah suara tersebut menyampaikan beberapa informasi di dalam kata yang diucapkan melalui gelombang suara. Informasi suara tersebut dapat diketahui melalui fitur suara itu sendiri, diantaranya dengan *pitch* dan *formant*, dimana *pitch* adalah frekuensi fundamental dari sinyal suara yang dihasilkan karena getaran pita suara, dan *formant* adalah frekuensi resonansi akustik dari bidang suara manusia. Kedua fitur tersebut merupakan fitur suara yang sangat penting untuk mengidentifikasi suara yang diucapkan dari seseorang.

Tahapan analisis pengidentifikasian suara ini dengan tahap *pre-processing*, *feature extraction*, dan *classification*. Tahapan *pre-processing* dengan *pre-emphasis*, *frame blocking*, dan *hamming window* dari sinyal suara. Tahapan *feature extraction* dengan *autocorrelation* untuk menentukan *pitch* dan *linear predictive coding* untuk menentukan *formant*, serta Tahapan *classification* dengan s*upport vector machine* untuk mengklasifikasikan fitur suara yang akan digunakan untuk mengidentifikasi suara tinggi dan rendah.

Berdasarkan hasil pengujian terhadap simulator dengan kesimpulan bahwa rata-rata nilai *pitch* pria untuk suara tinggi maupun rendah lebih rendah dari wanita, rata-rata nilai *formant* ke-1 dan *formant* ke-2 wanita untuk suara tinggi dan rendah lebih tinggi dari pria, rata-rata *formant* ke-3 pria untuk suara tinggi dan rendah lebih tinggi dari wanita, fitur suara yang lebih cocok merepresentasikan suara tinggi untuk pria dan wanita adalah *pitch*, kisaran *pitch* dan *formant* untuk suara pria adalah 191.1435 Hz (F0), 503.9955 Hz (F1), 956.9225 Hz (F2), dan 1561.43 Hz (F3). Untuk suara wanita adalah 268.51 Hz (F0), 534.871 Hz (F1), 1080.4475 Hz (F2), dan 1690.835 Hz (F3). Simulator mampu memprediksi suara tinggi dan rendah baik untuk pria dan wanita dengan persentase akurasi sebesar 70% untuk suara pria dan 100% untuk suara wanita.

Kata Kunci : *Pitch, Formant, Linear Predictive Coding, Support Vector Machine*.

1. PENDAHULUAN

Masalah yang dibahas dalam tugas akhir ini mengenai suara tinggi dan rendah manusia. Analisis ini dilakukan atas dasar untuk dapat mengidentifikasi suara tinggi dan rendah dengan melihat nilai-nilai fitur suara, yaitu *pitch* dan *formant*. Nilai *pitch* dan *formant* tersebut didapat melalui proses *feature extraction*. Nilai *pitch* dan *formant* suara tinggi dan rendah kemudian disimpan sebagai data pelatihan dan dibandingkan dengan nilai *pitch* dan *formant* yang ada pada data prediksi untuk diklasifikasikan dengan menggunakan metode *support vector machine* (SVM) agar diketahui apakah hasil klasifikasi nilai *pitch* dan *formant* tersebut masuk ke kategori tinggi atau rendah.

Hasil penelitian terhadap beberapa pustaka menjelaskan bahwa metode *support vector machine* merupakan metode klasifikasi jenis terpandu (*supervised*) yang bekerja menggunakan pemetaan *linear* dan *non-linier* dan sangat baik jika digunakan untuk mengklasifikasikan fitur yang mempunyai dua kelas atau *group*, dimana kelas atau *group* disini adalah suara tinggi dan rendah. Solusi yang dihasilkan dari metode *support vector machine* untuk penentuan suara tinggi adalah ketepatan pengidentifikasian suara tinggi dari fitur suara yang diuji, dimana fitur suara yang diuji adalah *pitch* dan *formant*.

1.1 *Autocorrelation*

Autocorrelation adalah korelasi silang terhadap sinyal dengan dirinya sendiri. Nilai autokorelasi suatu sinyal wicara akan menunjukan bagaimana gelombang suara itu membentuk suatu korelasi kepada dirinya sendiri. Bentuk-bentuk yang sama pada setiap delay waktu tertentu menunjukan perulangan bentuk pola sinyal suara. Dengan demikian akan bisa dilakukan estimasi nilai frekuensi fundamentalnya.

1.2 *Linear Predictive Coding*

Linear predictive coding merupakan salah satu metode pemodelan suara yang didasarkan pada teori bahwa bahwa sebuah sinyal suara manusia pada waktu n, sinyal suara dapat diperkirakan sebagai kombinasi linier dari p sinyal suara manusia sebelumnya. Tujuan dari metode lpc adalah untuk memisahkan efek *formant* dengan *pitch* atau frekuensi dasar manusia.

1.3 Support Vector Machine

Support vector machine (SVM) merupakan metode klasifikasi jenis terpandu (*supervised*) karena ketika pelatihan, diperlukan target pembelajaran tertentu. SVM dapat digunakan untuk klasifikasi yang dapat diterapkan pada deteksi tulisan tangan, pengenalan obyek, identifikasi suara dan lain-lain.

SVM adalah metode *machine learning* yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah *class* pada *input space.*

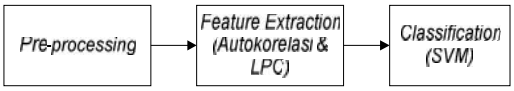
Berbeda dengan strategi *neural network* yang berusaha mencari *hyperplane* pemisah antar *class*, SVM berusaha menemukan *hyperplane* yang terbaik pada *input space.* Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik berfungsi sebagai pemisah dua buah *class* pada *input space.*

2 ISI PENELITIAN

2.1 Metode Pengumpulan Data

Tahap pengumpulan data yang digunakan pada penelitian ini adalah sebagai berikut :

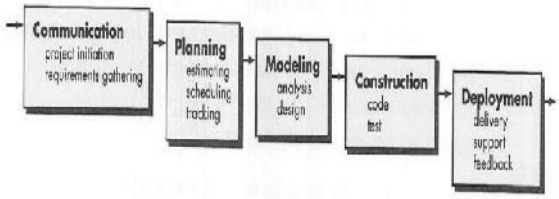
a. Studi Literatur
   Studi literatur adalah metode pengumpulan data yang diperoleh dengan mempelajari teori yang berkaitan dengan pengenalan suara, termasuk penerapan pemrosesan sinyal suara dan *support vector machine.*
b. Data dan Variabel Penelitian
   Data yang digunakan dalam penelitian ini adalah data tentang nilai frekuensi *pitch* dan *formant* dari suara pria dan wanita yang diperoleh dengan cara perekaman. Pada penelitian ini, hanya digunakan dua kategori (group), yaitu kategori suara tinggi dan suara rendah.
c. Langkah-langkah Analisis
   Langkah-langkah analisis identifikasi suara adalah sebagai berikut :
   1. Melakukan pengumpulan data suara pria dan wanita dengan kategori tinggi dan rendah.
   2. Melakukan prediksi klasifikasi kategori suara tinggi dan suara rendah menggunakan metode analisis *Support Vector Machine* (SVM) dengan langkah-langkah penyelesaian yang ditunjukkan pada gambar 1 di bawah ini :



Gambar 1 Tahapan Analisis Identifikasi Suara

2.2 Metode Pembangunan Perangkat lunak

Metode yang digunakan dalam pembuatan perangkat lunak ini menggunakan model *waterfall* seperti pada gambar 2 di bawah ini.



Gambar 2 Model *Waterfall*

Model ini adalah model klasik yang melakukan pendekatan secara sistematis, berurutan dalam membangun *software* berkat penurunan dari satu fase ke fase lainnya. Tahap dari model ini adalah sebagai berikut :

1. *Communication*
   Tahap *communication* merupakan analisis terhadap kebutuhan simulator dan tahap untuk mengadakan pengumpulan data suara manusia dengan melakukan perekaman suara pria dan wanita usia 20-35 tahun.
2. *Planning*
   Tahap *planning* merupakan lanjutan dari proses *communication (analysis requirement).* Tahap ini akan menghasilkan dokumen *user requirement* atau bisa dikatakan sebagai data yang berhubungan dengan keinginan *user* dalam rencana pembuatan simulator yang akan dilakukan. Spesifikasi kebutuhan perangkat lunak akan dibagi kedalam dua bagian yaitu SKPL-F (Spesifikasi Kebutuhan Perangkat Lunak Fungsional) dan SKPL-NF (Spesifikasi Kebutuhan Perangkat Lunak Non-Fungsional).
3. *Modeling*
   Tahap *modeling* akan menerjemahkan syarat kebutuhan ke sebuah perancangan simulator yang dapat diperkirakan sebelum dibuat *coding*. Proses ini berfokus pada rancangan detail (algoritma) prosedural. Tahapan ini akan menghasilkan dokumen yang disebut *software requirement.* Spesifikasi kebutuhan simulator dilakukan berdasarkan kebutuhan simulator pengidentifikasi suara untuk penerapan analisis algoritma serta hasil observasi. Spesifikasi kebutuhan perangkat lunak akan

dibagi kedalam dua bagian, yaitu SKPL-F (Spesifikasi Kebutuhan Perangkat Lunak Fungsional) dan SKPL-NF (Spesifikasi Kebutuhan Perangkat Lunak Non-Fungsional).

4. *Construction*
   Tahap *construction* merupakan proses pembuatan kode. *Coding* atau pengkodean merupakan penerjemahan desain dalam bahasa yang bisa dikenali oleh komputer. Tahapan inilah yang merupakan tahapan secara nyata dalam mengerjakan suatu simulator, artinya dalam tahapan ini penggunaan komputer akan dimaksimalkan. Setelah pengkodean selesai maka akan dilakukan *testing* terhadap simulator yang telah dibuat. Tujuan *testing* adalah menemukan kesalahan-kesalahan terhadap simulator tersebut untuk kemudian bisa diperbaiki.
5. *Deployment*
   Tahap *deployment* bisa dikatakan *final* dari pembuatan simulator. Setelah melakukan analisis, desain dan pengkodean, maka simulator yang sudah jadi akan digunakan oleh *user*. Kemudian *software* yang telah dibuat harus dilakukan pemeliharaan secara berkala. Pada simulator pengidentifikasian suara ini, tahap *deployment* tidak perlu dilakukan.

2.3 Analisis Metode

2.3.1 Pre-processing

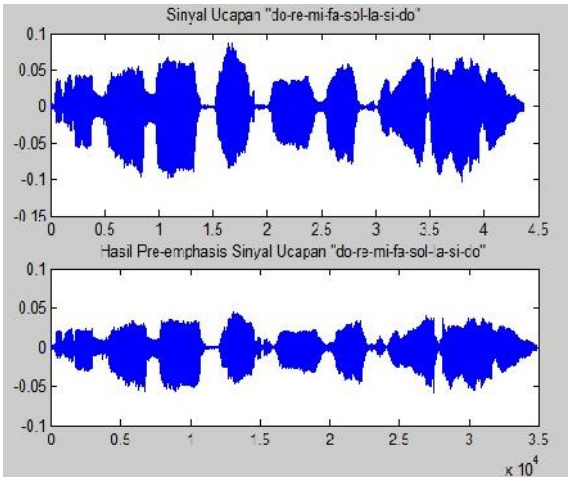
Pre-processing merupakan tahapan untuk mendapatkan data yang dibutuhkan untuk pengidentifikasian suara tinggi dan rendah. Pada tahap ini dilakukan akuisisi sampel sinyal suara dengan merekam suara responden dengan menggunakan perangkat lunak audacity, menghilangkan efek noise dengan *filter pre-emphasis*, menempatkan sinyal suara kedalam sejumlah *frame* dengan *frame blocking*, dan meminimalkan efek diskontinuitas pada potongan *frame* sinyal suara dengan *hamming window*.

a. Pre-emphasis

Pre-emphasis dilakukan untuk menghilangkan informasi yang tidak relevan dan kebisingan dengan menggunakan perhitungan *low pass filter* pada sinyal suara. *Pre-emphasis* merujuk pada proses memaksimakan kualitas sinyal analog dengan meminimalkan efek *noise* seperti distorsi selama perekaman dan transmisi data.

Pre-emphasis diperlukan dengan tujuan untuk mendapatkan bentuk *spectral* frekuensi sinyal suara yang lebih halus.

Sinyal suara hasil pre-emphasis dapat dilihat pada gambar 3.



Gambar 3 Sinyal Suara Hasil *Pre-emphasis*

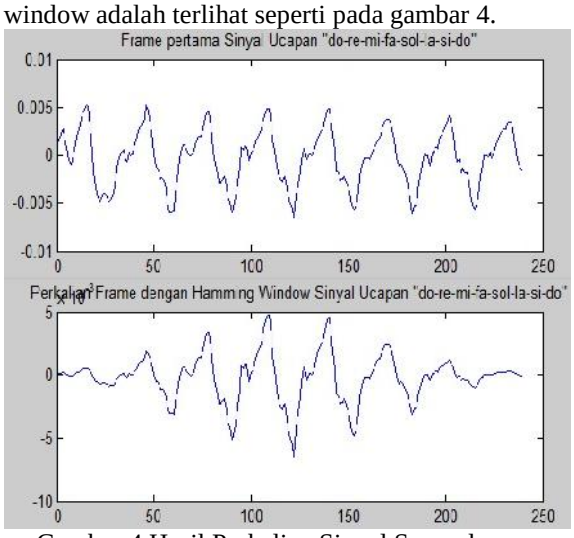
b. Frame Blocking

Sinyal suara hasil *pre-emphasis* kemudian ditempatkan kedalam *frame* menjadi beberapa bagian, dimana setiap frame sepanjang 30 milidetik dan dipisahkan sejauh 20 milidetik yang akan memudahkan dalam perhitungan dan analisa suara.

c. Hamming Window

Hamming window diperlukan untuk mengurangi efek diskontinuitas dari potongan - potongan sinyal suara yang berada pada setiap *frame*.

Sinyal suara hasil perkalian dengan hamming window adalah terlihat seperti pada gambar 4.



Gambar 4 Hasil Perkalian Sinyal Suara dengan Hamming Window

2.3.2 Feature Extraction

Feature extraction atau ekstraksi fitur adalah proses untuk mencari nilai fitur suara, dimana fitur suara yang diambil adalah *pitch* dan *formant*. Metode yang dipakai untuk mendapatkan nilai *pitch* adalah *autocorrelation*, sementara untuk mendapatkan nilai *formant* adalah *linear prediction coding*.

2.3.2.1 Pitch

Pitch merupakan frekuensi fundamental (F0) dari sinyal suara yang merupakan hasil akustik kecepatan getaran pita suara, semakin besar getaran pita suara, maka akan semakin tinggi nilai *pitch*. Periode *pitch* berkisar antara 10 sampai 20 milidetik. Setiap manusia mempunyai kisaran *pitch* tersendiri, tergantung dari pangkal tenggorok yang dimiliki. Kisaran *pitch* yang khas (*habitual pitch*) dimiliki oleh kebanyakan pria sebesar 50Hz - 250Hz, sedangkan wanita memiliki *pitch* (*habitual pitch*) yang lebih tinggi dibandingkan dengan pria, yaitu berkisar antara 120 - 500Hz. Frekuensi fundamental ini berubah secara konstan dan memberikan informasi *linguistic* seseorang seperti pembeda antara intonasi dan emosi.

Pada laki-laki ketika bersuara trakea dan laring pada tenggorokan membuka lebih lebar dibandingkan pada perempuan. Ukuran pita suara pada laki-laki berkisar antara 17.5 mm sampai 25 mm, sedangkan pada perempuan ukuran pita suaranya berkisar antara 12.5 sampai 17.5 mm. karena ukuran pita suara perempuan lebih kecil, maka suara yang dihasilkan oleh perempuan akan lebih tinggi.

2.3.2.2 Formant

Formant adalah frekuensi resonansi alami yang terjadi didalam rongga bidang suara, tergantung pada bentuk dan ukuran bidang suara dan lebih menyerupai gaung. Frekuensi *formant* bersifat tidak terbatas namun, untuk mengidentifikasi seseorang paling tidak ada 3 (tiga) format yang dianalisa yaitu, *formant* 1 (F1), *formant* 2 dan *formant* 3 (F3).

2.3.3 Classification

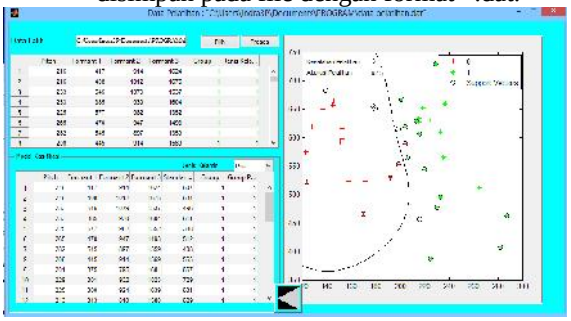
Classification atau klasifikasi adalah proses pengklasifikasian data fitur suara, dimana fitur suara dalam hal ini adalah *pitch* dan *formant* akan diklasifikasikan dengan metode klasifikasi *support vector machine* untuk menghasilkan informasi suara yang dihasilkan dari kedua fitur suara tersebut.

a. Pelatihan
   Pelatihan merupakan tahap yang dilakukan untuk membentuk data model klasifikasi suara, dimana data model ini akan digunakan sebagai data acuan klasifikasi suara dari data latih. Data latih yang digunakan dalam simulator pengideintifikasi suara ini merupakan data fitur suara dan *group* yang merupakan fitur suara tinggi dan rendah 40 responden yang terdiri dari 20 pria dan 20 wanita.
b. Prediksi
   Pada proses ini terlebih dahulu memasukan data yang akan diprediksi yaitu data *pitch* dan *formant*, kemudian dicocokan dengan model yang telah didapat dari hasil proses pelatihan.

2.3.4 Implementasi

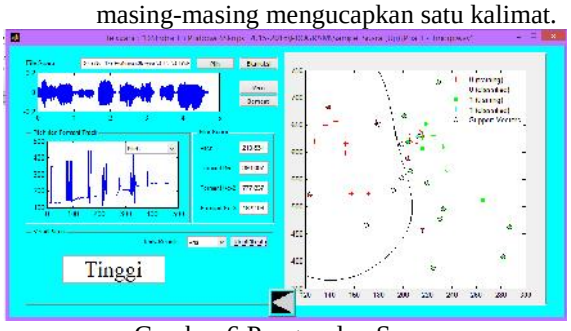
Berikut ini adalah implementasi dari aplikasi simulator pengidentifikasi suara.

a. Pengujian Pemodelan Klasifikasi
   Pemodelan klasifikasi merupakan tahap yang dilakukan untuk membentuk data model pelatihan klasifikasi suara, dimana data model ini akan digunakan sebagai data acuan klasifikasi suara dari data latih. Data latih yang digunakan dalam simulator pengideintifikasi suara ini merupakan data fitur suara dan *group* yang merupakan fitur suara tinggi dan rendah 40 responden yang terdiri dari 20 pria dan 20 wanita yang disimpan pada file dengan format *.dat.

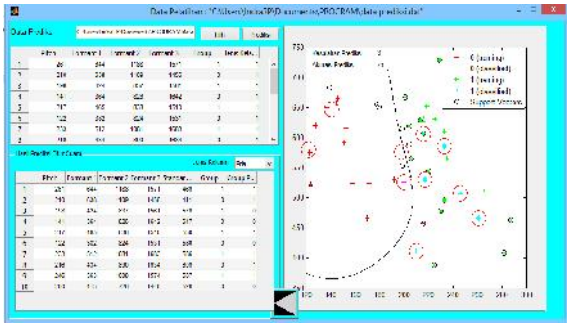


Gambar 5 Model Klasifikasi Suara

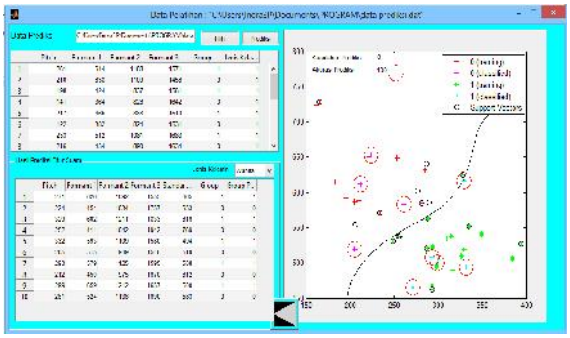
b. Pengujian Pengenalan Suara
   Pengujian pengenalan suara dilakukan dengan melakukan ekstraksi fitur suara dan prediksi suara terhadap file suara hasil perekaman suara responden yang berjumlah 10 orang, yang terdiri dari 5 responden pria, dan 5 responden wanita, dimana responden masing-masing mengucapkan satu kalimat.



Gambar 6 Pengenalan Suara



Gambar 7 Hasil Klasifikasi Fitur Suara Pria



Gambar 8 Hasil Klasifikasi Fitur Suara Wanita

Berdasarkan hasil pengujian terhadap perangkat lunak simulator dapat ditarik kesimpulan bahwa :

1. Rata-rata nilai *pitch* pria untuk suara tinggi maupun rendah lebih rendah dari wanita.
2. Rata-rata nilai *formant* ke-1 dan *formant* ke-2 wanita untuk suara tinggi dan rendah lebih tinggi dari pria.
3. Rata-rata *formant* ke-3 pria untuk suara tinggi dan rendah lebih tinggi dari wanita.
4. Fitur suara yang lebih cocok merepresentasikan suara tinggi untuk pria dan wanita adalah *pitch.*
5. Kisaran *pitch* dan *formant* untuk suara pria dan wanita adalah sebagai berikut :

   Pria :

   a. F0
      Minimal = 121.649 Hz
      Maksimal = 260.638 Hz
      Rata-rata = 191.1435 Hz
   b. F1
      Minimal = 363.971 Hz
      Maksimal = 644.02 Hz
      Rata-rata = 503.9955 Hz
   c. F2
      Minimal = 725.905 Hz
      Maksimal = 1187.94 Hz
      Rata-rata = 956.9225 Hz
   d. F3
      Minimal = 1440.13 Hz
      Maksimal = 1682.73 Hz
      Rata-rata = 1561.43 Hz

   Wanita :

   a. F0
      Minimal = 204.869 Hz
      Maksimal = 332.151 Hz
      Rata-rata = 268.51 Hz
   b. F1
      Minimal = 410.921 Hz
      Maksimal = 658.821 Hz
      Rata-rata = 534.871 Hz
   c. F2
      Minimal = 948.775 Hz
      Maksimal = 1212.12 Hz
      Rata-rata = 1080.4475 Hz
   d. F3
      Minimal = 1548.47 Hz
      Maksimal = 1833.2 Hz
      Rata-rata = 1690.835 Hz

Simulator mampu memprediksi suara tinggi dan rendah baik untuk pria dan wanita, serta perangkat lunak simulator yang dibangun telah cukup memenuhi tujuan awal pembangunan perangkat lunak simulator pengidentifikasi suara, ini dibuktikan dengan pelatihan data dan prediksi data yang menghasilkan persentase akurasi sebesar 70% untuk prediksi fitur suara pria dan 100% untuk prediksi suara wanita.

3 PENUTUP

Pada bab ini berisikan kesimpulan dari hasil penelitian yang telah dilakukan dan saran untuk perbaikan dan pengembangan penelitian lebih lanjut.

3.1 Kesimpulan

Dari hasil penelitian ini dapat ditarik kesimpulan mengenai aspek yang menjadi bahasan pada pengidentifikasi suara tinggi dan rendah, yaitu :

1. Simulator dapat mengidentifikasi suara tinggi dan rendah untuk pria dan wanita dengan menggunakan metode klasifikasi *support vector machine*, dengan ketepatan pelatihan dan prediksi yang baik.
2. Suara tinggi dan rendah wanita lebih mudah diidentifikasi dibandingkan dengan suara tinggi dan rendah pria, ini dibuktikan dengan pengujian sampel suara yang menghasilkan akurasi pengujian sebesar 100% wanita, dan 70% pria dengan fitur suara yang paling dominan mempengaruhi hasil prediksi suara tinggi dan rendah pada pria dan wanita adalah *pitch*.

3.2 Saran

Berdasarkan kesimpulan yang telah diuraikan, diharapkan simulator pengidentifikasi suara ini dikembangkan lebih baik lagi, agar suara yang dikenali lebih banyak.

DAFTAR PUSTAKA

[1] Endah, Sukmawati Nur dan Dinar Mutiara. 2012. *Analisis Pitch dan Formant Sinyal Ucapan Kata*. Prosiding Seminar Nasional Ilmu komputer. Semarang.

[2] Fadlisyah, Bustami, dan Ikhwanus, M, 2013. *Pengolahan Suara*. Yogyakarta: Graha Ilmu.

[3] Wicaksono, Galieh., Prayudi, Yudi. *Teknik Forensika Audio Untuk Analisa Suara Pada Barang Bukti Digital*. Universitas Islam Indonesia, Yogyakarta: Pusat Studi Forensika Digital.

[4] Pudjo, Widodo, Prabowo., Trias, Handayanto, Rahmadya., Herlawati. 2013, *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains.

[5] Pan, Yixiong., Shen, Peipei., and Shen,

Liping. 2012, *Speech Emotion Recognition Using Support Vector Machine*, International Journal of Smart Home (IJSH).

[6] Roger S. Pressman. 2010. *Software Engineering: A Practitioner's Approach*, 4th ed. New York: McGraw-Hill Companies.

[7] Bagas, Bhaskoro, Susetyo., Riedho, Altedzar. 2012, *Aplikasi Pengenalan Gender Menggunakan Suara*. Seminar Nasional Aplikasi Teknologi Informasi.

[8] Suyanto, S.T., Msc. 2007. *Artificial Intelegence*. Bandung: Informatika.

[9] Rabiner, Lawrence., Juang, Bing-Hwang. 1993, *Fundamental of Speech Recognition*. New Jersey: Prentince-Hall.

[10] Rabiner, Lawrence., W, Schafer, Ronald. 1978, *Digital Processing of Speech Signals*. New Jersey: Prentice-Hall, Inc.

[11] Hadi, Putra, Prabowo. *Penggolongan Suara Berdasarkan Usia dengan Menggunakan Metode K-Means*. Surabaya: Institut Teknologi Sepuluh Nopember.

[12] C, Snell, Roy., Milinazzo, Fausto. 1993, Formant Location From LPC Analysis Data. *IEEE Transaction on Speech and Audio Processing*, vol. I.

[13] Suyanto. 2011, *Artificial Intelligence*. Bandung, Indonesia: Informatika.

[14] Alpaydın, Ethem. 2010, *Introduction to Machine Learning Second Edition*. London: The MIT Press.

[15] Bernhard E. Boser and Isabelle M. Guyon and Vladimir Vapnik. 1992, *A Training Algorithm for Optimal Margin Classifiers*. Proceedings of the fifth annual workshop on Computational learning theory (COLT).

[16] Ivanciuc, Ovidiu. 2007, *Applications of Support Vector Machines in Chemistry*. Reviews in Computational Chemistry, vol. 23.

[17] Hermawati, Astuti, Fajar. 2013. *Data Mining*. Yogyakarta: Andi.

[18] Kantardzk, Mehmed. 2011. *Data Mining (Concepts, Models, Methods, and Algorithms)*. New Jersey: John Wiley & Sons.

[19] McKinney, James (1994). The Diagnosis and Correction of Vocal Faults. Genovex Music Group. ISBN 978-1-56593-940-0.