

Identifikasi Suara Vokal Suku Banjar Berdasarkan Frekuensi Formant

Arfan Eko Fahrudin¹⁾, Nofida Risna Diyanti²⁾ dan Tetti Novalina Manik¹⁾

Abstrak: Telah dibuat program untuk mengidentifikasi suara vokal suku Banjar berdasarkan frekuensi formant dengan menggunakan metode LPC (*Linear Predictive Coding*). Sampel suara vokal diambil dari 30 responden yang terdiri dari 15 perempuan dan 15 laki-laki untuk setiap suara vokal dengan frekuensi *sampling* 8000 Hz. Proses yang dilakukan dalam identifikasi suara vokal suku Banjar antara lain: *sampling, pre emphasis filter, frame blocking, windowing* dan ekstraksi ciri dengan menggunakan metode LPC. Hasil frekuensi formant rata-rata untuk suara vokal perempuan adalah sebagai berikut; untuk suara vokal /a/ F1: 861.96 Hz, F2 : 1227.17 Hz, F3: 1686.92 Hz, suara vokal /e/ F1: 783.34 Hz, F2: 1365.78 Hz, F3: 2207.58 Hz, suara vokal /i/ F1: 446.94 Hz, F2: 1288.10 Hz, F3: 2583.38 Hz, suara vokal /o/ F1: 780.75 Hz, F2:1064.41 Hz, F3: 1536.06 Hz, suara vokal /u/ F1: 523.71 Hz, F2: 920.18 Hz, F3:1789.06 Hz. Hasil frekuensi formant rata-rata untuk suara vokal laki-laki adalah sebagai berikut; untuk suara vokal /a/ F1: 891.09 Hz, F2: 1338.58 Hz, F3: 1844.57 Hz, suara vokal /e/ F1: 620.28 Hz, F2: 1446.42 Hz, F3: 1907.88 Hz, suara vokal /i/ F1: 391.00 Hz, F2: 1490.83 Hz, F3: 2313.92 Hz, suara vokal /o/ F1: 663.38 Hz, F2: 977.77 Hz, F3: 2124.76 Hz, suara vokal /u/ F1: 488.91 Hz, F2: 912.18 Hz, F3: 2074.57 Hz. Hasil identifikasi menunjukkan bahwa setiap suara vokal /a/, /e/, /i/, /o/ dan /u/ mempunyai karakteristik frekuensi formant F1, F2 dan F3 yang khas, sehingga dapat digunakan sebagai pembeda antara vokal satu dengan yang lain.

Kata Kunci: *vokal suku Banjar, frekuensi formant, LPC (Linear Predictive Coding)*

PENDAHULUAN

Ucapan manusia dihasilkan oleh suatu sistem produksi ucapan yang dibentuk oleh alat-alat ucap manusia. Proses tersebut dimulai dengan formulasi pesan dalam otak pembicara. Pesan tersebut akan diubah menjadi perintah-perintah yang diberikan kepada alat-alat ucap manusia, sehingga akhirnya dihasilkan ucapan yang sesuai dengan pesan yang ingin diucapkan [2].

Proses pembentukan ucapan manusia melibatkan tiga proses utama, yaitu pembangkitan sumber (*source*

generator), artikulasi (*articulation*), serta radiasi (*radiation*). Proses pembentukan bunyi bahasa dimulai dengan memanfaatkan pernapasan sebagai sumber tenaganya. Pada saat manusia mengeluarkan nafas, paru-paru manusia akan menghembuskan tenaga berupa arus udara. Arus udara dapat mengalami perubahan pada pita suara yang terletak pada pangkal tenggorokan. Arus udara dari paru-paru itu dapat membuka ke dua pita suara yang rapat sehingga menyebabkan corak bunyi tertentu. Gerakan

¹⁾ Staf Pengajar PS Fisika FMIPA Universitas Lambung Mangkurat, Banjarbaru

²⁾ Mahasiswa PS Fisika FMIPA Universitas Lambung Mangkurat, Banjarbaru

membuka dan menutup pita suara itu akan mengakibatkan arus udara dan udara yang berada di sekitar pita suara akan turun berubah tekanan dan ikut bergetar [5].

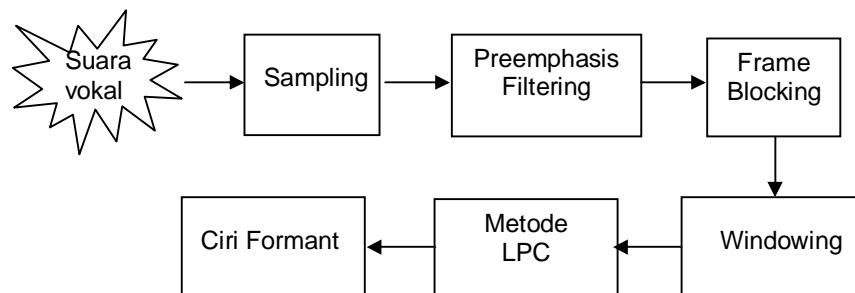
Vokal dan konsonan adalah elemen dasar dari tiap bahasa. Selain itu perbedaan pengucapan suatu kata dari orang dengan bahasa ibu yang berbeda disebabkan oleh variasi vokal dan cara pengucapannya [9]. Ketika berbicara, vokal dihasilkan akibat getaran pada pita suara tanpa adanya penyempitan pada saluran vokal (*vokal tract*). Jenis suara vokal yang dihasilkan dipengaruhi oleh posisi lidah, rahang, serta bibir. Suara vokal biasanya mempunyai durasi yang lebih lama dalam pengucapannya dibandingkan konsonan, sehingga sangat mudah dan memungkinkan untuk dikenali bunyi pengucapannya [6].

Pengenalan suara vokal salah satunya dilakukan berdasarkan frekuensi formant masing-masing vokal [3][7]. Frekuensi formant merupakan frekuensi resonansi pada *vocal tract*

(saluran suara) ketika seseorang mengucapkan suara vokal [9]. Frekuensi yang dominan dari frekuensi formant adalah tiga frekuensi pertamanya, biasanya dituliskan dengan notasi F1, F2 dan F3 [4]. Metode yang digunakan untuk menentukan frekuensi formant dari suara vokal diantaranya dengan metode *Linear Predictive Coding* (LPC) [3][8], Hidden Markov Models [4], cepstrum [3] dan metode *autoregressive* [7]. Dalam penelitian ini dilakukan identifikasi suara vokal suku Banjar berdasarkan frekuensi formant dengan menggunakan metode LPC.

METODOLOGI PENELITIAN

Data suara vokal input berupa vokal /a/, /e/, /i/, /o/ dan /u/ diperoleh dari sampel 30 orang yang terdiri dari 15 laki-laki dan 15 perempuan dewasa dari suku Banjar. Pengambilan data dilakukan dengan merekam suara vokal menggunakan *microphone* yang tersambung komputer yang dilengkapi dengan *software* Audacity 1.3 Beta.



Gambar 1. Blok diagram penelitian

Proses identifikasi suara vokal bahasa Indonesia untuk suku Banjar yang dibuat dalam penelitian ini adalah seperti terlihat pada gambar 1.

1. *Sampling*

Pada proses pencuplikan (*sampling*), sinyal analog $x(t)$ dirubah menjadi sinyal diskrit $x_i = x(iT)$ dengan waktu periodik $t_i = iT$ (i adalah bilangan integer). Disini $T[s]$ merupakan periode *sampling*, sehingga $S = 1/T [Hz]$ adalah frekuensi *sampling* [8]. Frekuensi *sampling* yang digunakan dalam penelitian ini adalah sebesar 8 KHz, ini untuk memenuhi kriteria Nyquist yang menyatakan bahwa frekuensi *sampling* minimal adalah dua kali dari frekuensi sinyal yang *disampling* (sinyal wicara (*speech*) berada pada daerah frekuensi 300-3400 Hz).

2. *Preemphasis Filtering*

Filter pre-emphasis didasari oleh hubungan input/output dalam domain waktu yang dinyatakan dalam persamaan berikut [6]:

$$\bar{s}(n) = s(n) - a s(n-1) \quad \dots(1)$$

Dengan a merupakan konstanta filter pre-emphasis, biasanya bernilai $0.9 \leq a \leq 1$. Dalam bentuk dasar operator z sebagai unit filter, persamaan diatas akan memberikan sebuah *transfer function* filter pre-emphasis seperti berikut:

$$H(z) = 1 - a z^{-1} \quad \dots(2)$$

3. *Frame blocking*.

Proses *frame blocking* yaitu melakukan blok terhadap sinyal-sinyal menjadi frame-frame N sampel, dengan frame-frame berdekatan dengan spasi M ($M < N$). Frame pertama terdiri dari N sampel pertama. Frame kedua dengan M sampel setelah frame pertama, dan overlap dengan $N-M$ sampel. Dengan cara yang sama, frame ketiga dimulai $2M$ sampel setelah frame pertama (atau M sampel setelah frame kedua) dan overlap dengan $N-2M$ sampel. Proses ini berlanjut hingga semua sinyal suara dihitung dalam satu atau banyak frame. Sinyal baru yang diperoleh adalah sebagai berikut:

Jadi, proses frame tersebut dilakukan secara terus-menerus hingga semua sinyal dapat terproses. Selain itu, proses ini umumnya dilakukan secara overlapping untuk setiap frame-nya. Panjang daerah overlap yang umum digunakan adalah kurang lebih 30% sampai 50% dari panjang frame.

4. *Windowing*

Tahap selanjutnya berupa proses penjedeleaan (*windowing*) terhadap tiap frame hasil *frame blocking* yang bertujuan untuk meminimalkan diskontinuitas sinyal pada bagian awal dan akhir tiap frame [6]. Hasil dari sinyal ter*window* adalah :

$$x(n) = xl(n) \cdot w(n), \quad 0 \leq n \leq N - 1 \quad \dots(3)$$

$$x(n) = xl(n) \cdot w(n), \quad 0 \leq n \leq N - 1 \quad \dots(4)$$

dengan $w(n)$ adalah window hamming yang dinyatakan dengan persamaan:

5. Ekstraksi ciri dengan metode LPC

Sinyal suara dapat dimodelkan dengan persamaan sebagai berikut [3]:

$$s(n) = - \sum_{i=1}^{N_{LP}} a_{LP}(i) \cdot s(n-i) + e(n) \quad \dots(5)$$

dengan N_{LP} , a_{LP} dan $e(n)$ masing-masing adalah jumlah koefisien dalam model, koefisien prediksi linier dan

error dari model. Persamaan (5) dapat dituliskan dalam notasi transformasi Z sebagai operasi pemfilteran linier:

$$S(z) = - \sum_{i=1}^{N_{LP}} a_{LP}(i) z^{-i} \cdot S(z) + E(z) \quad \dots(6)$$

sehingga $E(z) = H_{LP}(z) \cdot S(z)$ $E(z)$ dan $S(z)$ masing-masing merupakan transformasi Z dari sinyal

error dan sinyal bicara. $H_{LP}(z)$ didefinisikan sebagai invers filter prediksi linier :

$$H_{LP}(z) = \sum_{i=1}^{N_{LP}} a_{LP}(i) z^{-i} \quad \dots(7)$$

Frekuensi formant dapat diestimasi dari prediksi linier spectrum yang telah dihaluskan. Dari spectrum ini, nilai puncak maksima lokal yang merupakan frekuensi formant dapat diperoleh.

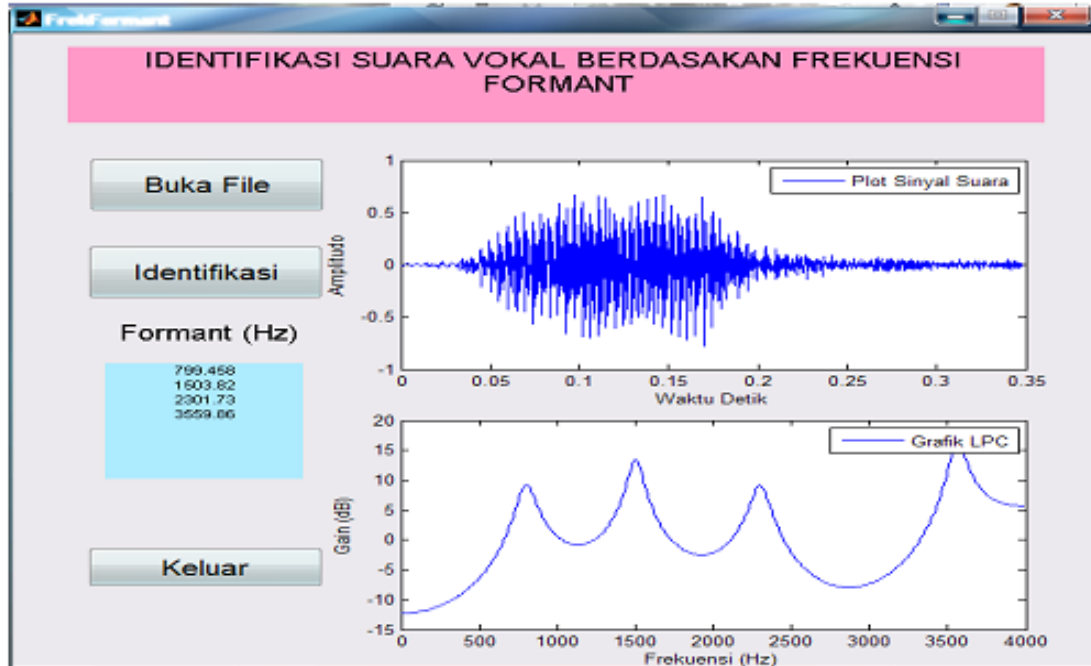
HASIL DAN PEMBAHASAN

Program yang digunakan untuk mengidentifikasi suara vokal dari sampel yang diperoleh dibuat dengan

Matlab 7.7, dengan tampilan *Graphical User Interface* (GUI) seperti pada Gambar 2. Dari gambar ditunjukkan grafik dari sinyal suara yang diidentifikasi (grafik bagian atas) serta grafik spektrum LPC yang sudah dihaluskan (grafik bagian bawah). Dari grafik spektrum LPC ditunjukkan puncak (*peak*) yang merupakan frekuensi formant yang dicari, puncak

yang pertama menunjukkan frekuensi formant 1 (F1), demikian pula untuk puncak ke-2 menunjukkan frekuensi

formant 2 (F2) dan seterusnya. Dalam penelitian ini digunakan tiga frekuensi pertama dari frekuensi formant.



Gambar 2. Tampilan GUI program

Karakteristik setiap vokal untuk masing-masing responden ditentukan dengan cara membandingkan frekuensi formant yang diperoleh tersebut. Data

frekuensi vokal yang diperoleh dari semua responden berdasarkan jenis kelaminnya dapat dilihat pada tabel berikut:

Tabel 1. Nilai rata-rata frekuensi Formant pada suara vokal perempuan

Vokal	F1 (Hz)	F2 (Hz)	F3 (Hz)
a	861.96	1227.17	1686.92
e	783.34	1365.78	2207.58
i	446.94	1288.1	2583.38
o	780.75	1064.41	1536.06
u	523.71	920.18	1789.06

Tabel 2. Nilai rata-rata frekuensi Formant pada suara vokal laki-laki

Vokal	F1 (Hz)	F2 (Hz)	F3 (Hz)
a	891.09	1338.58	1844.57
e	620.28	1446.42	1907.88
i	391	1490.83	2313.92
o	663.38	977.77	2124.76
u	488.91	912.18	2074.57

Berdasarkan data frekuensi formant yang diperoleh dari Tabel 1 dan Tabel 2 dapat diketahui bahwa frekuensi formant tiap vokal baik itu F1, F2 dan F3 mempunyai selisih frekuensi yang tidak berbeda jauh antara suara vokal laki-laki dan perempuan. Dari Tabel juga ditunjukkan bahwa setiap vokal mempunyai karakteristik F1, F2 dan F3 yang berbeda satu sama lain (mempunyai frekuensi formant yang khas), sehingga untuk membedakan tiap suara vokal tersebut dapat dimanfaatkan frekuensi formant sebagai ciri pengidentifikasinya.

KESIMPULAN

Hasil penelitian yang telah dilakukan diperoleh kesimpulan sebagai berikut:

1. Dengan metode LPC dapat diperoleh frekuensi formant suara vokal berdasarkan grafik spektrum LPC.

2. Setiap suara vokal /a/, /e/, /i/, /o/ dan /u/ mempunyai karakteristik frekuensi formant F1, F2 dan F3 yang khas, sehingga dapat digunakan sebagai pembeda antara vokal satu dengan yang lain.

DAFTAR PUSTAKA

- [1] Arfan Eko F., "Identifikasi Suara Vokal Berdasarkan Frekuensi Formant Dengan Metode Autoregressive", Proceeding SEMIRATA 2011, 2011.
- [2] Arman, A.A., "Proses Pembentukan dan Karakteristik Sinyal Ucapan", Departemen Teknik Elektro ITB, 2010.
- [3] Gargouri, D., Kammoun, M.A., Hamida, A., "A Comparative Study of Formant Frequencies Estimation Techniques", Proceeding of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, 2006.
- [4] Messaoud, Z., Gargouri, D., Zribi, Z., Hamida, A., "Formant Tracking Linear Prediction Model using HMMs for Noisy Speech Processing", International Journal of Information and Communication Engineering, 2009.

- [5] Nurlaily, "Pencocokan Pola Suara dengan Algoriyma FFT dan DC", Jurnal MIPA, Medan, 2009.
- [6] Rabiner, L., Bing-Hang Juang, "Digital Processing of Speech Signals", Prentice-Hall International Inc, 1993.
- [7] S.A.Mohd Yusof, P.M.Raj, S. Yaacob, "Speech Recognition Application based on Malaysian spoken vowels using Autoregressive Model of the Vokal Tract", Proceeding of International Conference On Electrical Engineering And Informatics, Institut Teknologi Bandung, 2007.
- [8] Sadaoki Furui, "Digital Speech Processing, Synthesis, and Recognition", Marcel Dekker Inc, 1989.
- [9] Sadegi, V.S., Yaghmaie K., "Vowel Recognition Using Neural Network", International Journal Of Computer Science And Network Society, VOL.6 No.12, December 2006.
- [10] Snell, R.C., Milinazzo, F., "Formant Location From LPC Analysis Data", IEEE Transactions On Speech and Audio Processing, Vol 1, No.2, April, 1993.
- [11] Titze, I.R., "Principles of Voice Production", Prentice Hall, 1994.