

Aplikasi Pengolahan Sinyal Suara pada Teknologi Kecerdasan Buatan

Djadjat Sudaradjat¹, Suryanto², Andi Rosano³

^{1,2,3}Universitas Bina Sarana Informatika
e-mail: 1djadjat.dsj@bsi.ac.id, 2suryanto.syt@bsi.ac.id, 3andi.aox@bsi.ac.id

Abstrak - Dalam konsep Kecerdasan Buatan (Artificial Intelligence), mesin dibuat untuk bisa bekerja dan berperilaku seperti manusia dengan menerapkan prinsip bahwa setiap aspek kecerdasan dapat dipelajari dan diwujudkan dalam suatu sistem. Salah satu alat komunikasi penting yang ada pada manusia adalah dengan menggunakan suara manusia atau ucapan untuk berbicara atau berkomunikasi dengan lawan bicaranya.. Pada suara manusia agar dapat dipelajari dan diwujudkan dalam suatu system, maka sinyal suara manusia perlu diolah dan diekstraksi menjadi parameter-parameter suara manusia. Karakter suara manusia yang unik dan beragam bila diolah dan diekstraksi akan memiliki parameter-parameter yang unik dan beragam pula. Ada beberapa aplikasi berbasis kecerdasan buatan yang mendukung system alat bicara ini diantaranya sistem Pengenal Ucapan (Speech Recognition) dan sistem Pengenal Pengucap/Pembicara (Speaker Recognition). Salah satu metoda yang akan diteliti dalam paper ini adalah metoda LPC (Linear Predictive Coding). Dengan metoda LPC maka akan dihasilkan parameter-parameter suara manusia sesuai dengan bentuk kata yang diucapkan dari beragam pembicara yang berlainan. Parameter-parameter suara manusia tersebut dapat dipelajari dan diwujudkan dalam suatu system Pengenal Ucapan sehingga dapat digunakan untuk alat kontrol atau mengendalikan suatu perangkat, misalnya perintah menghidupkan atau mematikan peralatan elektronik dan perintah-perintah lainnya. Aplikasi lainnya adalah sistem Pengenal Pembicara yang dapat digunakan untuk alat identifikasi pengguna, misalnya pada sistem password untuk membukanya menggunakan suara pengguna yaitu dengan cara mempelajari parameter-parameter suara pengguna yang unik sehingga system password dapat mengenalinya dan dapat membuka password.

Kata Kunci: Pengolahan sinyal suara, LPC, ANN, AI

Abstract. *In the concept of Artificial Intelligence (Artificial Intelligence), machines are made to be able to work and behave like humans by applying the principle that every aspect of intelligence can be learned and realized in a system. One of the important communication tools that exist in humans is to use human voice or speech to speak or communicate with interlocutors. In the human voice so that it can be learned and realized in a system, then the human voice signal needs to be processed and extracted into sound parameters human. The unique and diverse human voice character when processed and extracted will have unique and diverse parameters. There are several artificial intelligence based applications that support this speech system including Speech Recognition systems and Speaker Recognition systems. One method that will be examined in this paper is the LPC (Linear Predictive Coding) method. With the LPC method, parameters of human speech will be generated in accordance with the spoken form of a variety of different speakers. The parameters of the human voice can be learned and realized in a Speech Recognition system so that it can be used to control or control a device, for example commands to turn on or turn off electronic equipment and other commands. Another application is the Speaker Recognition system that can be used for user identification tools, for example in the password system to open it using the user's voice by learning unique user voice parameters so that the system password can recognize it and can open passwords.*

Keyword: *Speech Signal Processing, LPC, ANN, AI*

PENDAHULUAN

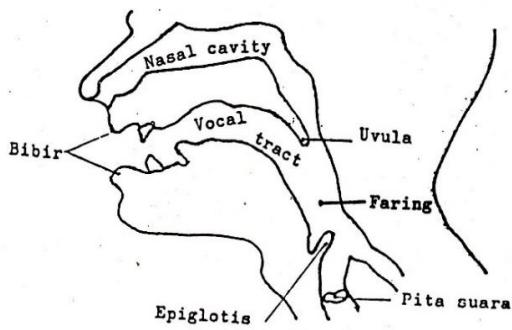
Kecerdasan buatan (AI) dikembangkan dengan cara meniru kecerdasan manusia yang kecerdasannya berkembang dengan cara belajar atau training. Pada makalah ini akan dibahas kecerdasan buatan pada mesin yang dapat mengenali suara manusia dengan cara belajar mengenali suara manusia (*Machine Learning*) sehingga antara mesin dan manusia dapat berkomunikasi (*Man Machine Interaction*). Algoritma komputer yang akan digunakan untuk meniru proses terbentuknya suara

manusia pada makalah ini menggunakan algoritma LPC (*Linear Predictive Coding*) dan algoritma komputer untuk dapat mengenali suara manusia salah satunya menggunakan algoritma ANN (*Artificial Neural Network*) meniru proses yang terjadi pada jaringan syaraf manusia.

Sinyal suara manusia membawa informasi tata Bahasa, nada pembicara, dan emosi pembicara. Pertukaran informasi suara memegang peranan yang sangat penting dalam kehidupan kita. Struktur tata Bahasa dan akustik dari suara manusia menunjukkan kemampuan intelektual kita, selain itu sangat erat

hubungannya dengan perkembangan social dan budaya kita. (Muslich, 1990) (Jurafsky, 2018)

Proses terbentuknya suara manusia dapat dijelaskan sebagai berikut; tiga faktor utama yang terlibat dalam pembentukan ucapan manusia adalah sumber tenaga, alat ucapan yang menimbulkan getaran, dan rongga pengubah getaran. Udara pernapasan dari paru-paru, sebagai sumber tenaga, dapat keluar melalui rongga mulut dan menghasilkan bunyi oral. Udara juga bisa keluar melalui hidung dan ini menghasilkan bunyi nasal (sengau). Bisa juga, arus udaranya keluar melewati hidung dan mulut; bunyi ini disebut bunyi dinasialisasi (bunyi yang disengaukan). Bagan di gambar 1 memperlihatkan semua alat ucapan yang “berjasa” dalam pembentukan bunyi Bahasa.



Gambar 1. Bagan mekanisasi pembentukan ucapan manusia.

Bunyi Bahasa yang dihasilkan oleh alat-alat ucapan ini dipelajari, salah satunya, dalam bidang fonetik artikulatoris. Kita tahu bahwa ketika paru-paru menghembuskan udara, pita suara dapat merapat dan merenggang. Pita suara merapat, maka bunyi Bahasa yang keluar terasa “berat”; ini menghasilkan bunyi tak-bergetar (unvoiced), contohnya [s]. Bila pita suara merenggang, arus udara mudah lewat dan bunyi dihasilkan dengan “ringan”; ini menghasilkan bunyi bergetar (voiced), misalnya [a]. Setelah melewati rongga faring, arus udara mengalir kebagian atas tenggorokan. Jika uvula menutup saluran ke rongga hidung, maka arus udara akan lewat mulut. Ini menghasilkan bunyi-bunyi seperti [p], [g], dan [f]. Bila ingin nasal, uvula diturunkan sampai menempel ke belakang lidah, dan udara bebas lewat hidung. Bunyi yang dihasilkan, misalnya [m], dan [n].

Dari mekanisme pembentukan ucapan tersebut diketahui parameter-parameter yang diperlukan untuk membentuk suatu ucapan. Model mekanisme pembentukan ucapan dengan parameter-parameternya dibuat sebagai model dibagian pengenal ucapan, dan dinamakan model sintesis. Di bagian ekstraksi sinyal ucapan, parameter-parameter untuk membentuk suatu ucapan tersebut dibangkitkan dengan cara menganalisis sinyal masukan, dan modelnya dinamakan model analisis.

Dari ketiga faktor utama yang terlibat dalam pembentukan suatu ucapan seperti alat ucapan yang menimbulkan getaran, sumber tenaga, dan rongga pengubah getaran dapat dibuat modelnya seperti terlihat di Gambar 2(b). Pita suara yang menimbulkan getaran dimodelkan sebagai dua jenis fungsi eksitasi, yaitu pertama jika bunyi yang dihasilkan adalah bergetar maka digunakan model pembangkit pulsa yang akan bekerja dengan periode sebesar periode pitch dan kedua jika bunyi yang dihasilkan adalah tak-bergetar maka digunakan model pembangkit desis (acak) yang akan bekerja. Sumber tenaga yang akan menentukan keras lembutnya bunyi dimodelkan sebagai penguatan (gain), sedangkan rongga pengubah getaran dimodelkan sebagai filter.

Karena filter yang digunakan pada model sintesis bersifat all-pole, maka model tersebut disebut model all-pole atau model AR (autoregressive) (Rabiner L. R., 2007). Filter tersebut diimplementasikan pada komputer dalam kawasan-z, dari Gambar 2 (b) diperoleh:

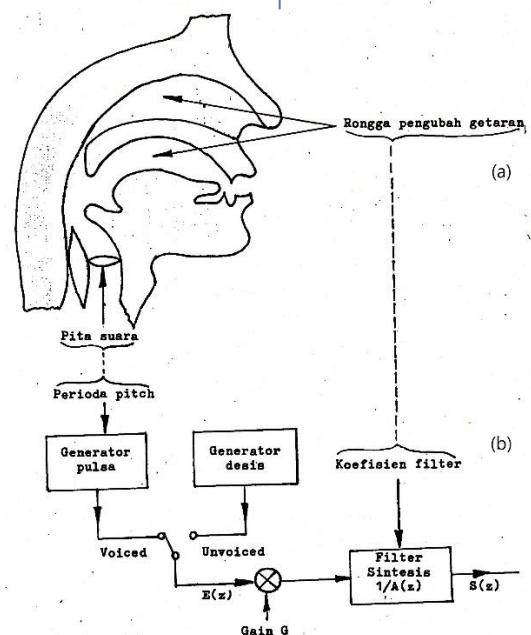
$$S(z) = E(z) \cdot G \cdot \frac{1}{A(z)} \quad (\text{model sintesis}) \dots \dots \dots (1)$$

dengan:

$$A(z) = \sum_{i=0}^p a(i)z^i$$

$$(a) \text{ koef. filter, } a(0) \neq 1 \dots \dots \dots (2)$$

$$S(z) \xrightarrow[Z]{Z^{-1}} s(nT) = s(t) \quad | \quad t = nT \dots \dots \dots (3)$$



Gambar 2. (a) Bagan pembentukan ucapan manusia, (b) Gambar model pembentukan ucapan manusia (model sintesis).

Sinyal waktu diskrit $s(nT)$ merupakan invers transformasi-z dari $S(z)$. Untuk periode waktu diskrit yang dinormalisasi $T=1$, maka :

$$S(z) \Leftrightarrow s(n).$$

Dengan hubungan yang sama dengan sinyal $s(n)$ diatas,

$$E(z) \Leftrightarrow e(n)$$

adalah model dari keluaran pita suara. Jika jenis bunyi bergetar, $e(n)$ akan berbentuk fungsi pulsa. Jika jenis tak-bergetar, $e(n)$ akan berbentuk fungsi acak. Sinyal $e(n)$ sesudah diperkuat G dilewatkan kedalam model rongga pengubah getaran dalam bentuk filter all-pole $1/A(z)$ untuk menghasilkan sinyal $s(n)$.

Dari persamaan (1) dapat diperoleh model analisis dalam bentuk persamaan :

$$E(z) = 1/G \cdot S(z) \cdot A(z) \text{ (model analysis)} \dots \dots \dots (4)$$

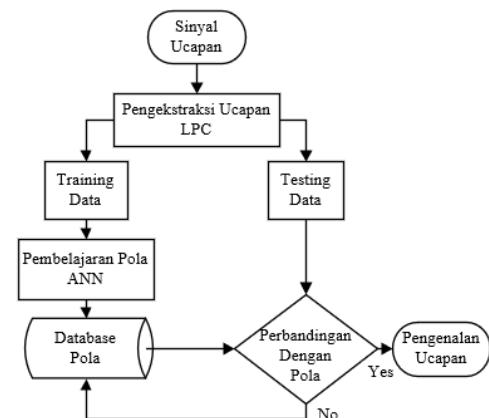
dimana $S(z)$ adalah transformasi-z dari sinyal suara $s(n)$ sebagai masukan invers filter (filter all-zero) $A(z)$, keluaran $E(z)$ akan membawa informasi parameter-parameter sinyal suara masukan tersebut.

Untuk pembentukan kembali suatu ucapan pada model sintesis diperlukan parameter-parameter berupa koefisien filter, gain, periode pitch P dan jenis bunyi voiced/unvoiced. Dengan demikian dibagian model analisis pada simulasi algoritma LPC akan dibagi dua algoritma perhitungan, yaitu pertama “algoritma analysis LPC” yang berisi perhitungan koefisien filter dan gain, kedua “algoritma periode pitch” yang berisi perhitungan periode pitch dan penentuan jenis bunyi bergetar dan tak-bergetar dari sinyal suara. (Kala, 2015)

Parameter-parameter LPC pembentukan suara manusia yang dihasilkan oleh tiap orang akan berbeda dan unik, sehingga dapat digunakan oleh komputer untuk mengenali suara manusia. Setiap sinyal suara yang diterima komputer untuk dikenali akan diproses oleh sistem LPC, dan parameter-parameter LPC yang dihasilkan akan disimpan dalam memori komputer untuk dikodekan kedalam bentuk digital. Data-data digital yang tersimpan dalam memori tersebut akan digunakan oleh komputer untuk proses pengenalan suara manusia yang diterimanya, dan setiap suara yang diterima oleh komputer akan dibandingkan dengan data suara yang ada di memori untuk disamakan artinya sehingga komputer akan mengenali suara yang diterimanya.

Jaringan syaraf tiruan merupakan salah satu sistem pengenal ucapan pada komputer yang dirancang meniru cara kerja otak manusia untuk menyelesaikan suatu masalah dengan melakukan proses belajar melalui perubahan bobot pada sinapsisnya (Siang., 2005). Jaringan syaraf tiruan mampu mengenali sinyal ucapan dengan cara mempelajari parameter-parameter ucapan dan membandingkannya dengan pola ucapan yang sudah pernah dipelajari sebelumnya, jika tidak ada pola yang sesuai, maka pola yang baru dipelajarinya akan disimpan di dalam memori sebagai basis data baru.

Blok diagram pembelajaran pola dan pengenalan ucapan ditunjukkan pada gambar 3.



Gambar 3. Blok diagram Pembelajaran Pola dan Pengenalan Ucapan.

Pada sistem pembelajaran pola dan pengenalan ucapan seperti yang ditunjukkan di gambar 3 merupakan sistem kecerdasan buatan AI yang meniru cara belajar dan mengenali ucapan pada manusia, setiap sinyal ucapan yang masuk pada sistem LPC akan mengekstraksi ucapan, data hasil ekstraksi akan dipelajari polanya oleh ANN dan disimpan di database pola, kemudian data yang akan diuji untuk dikenali dibandingkan dengan data yang ada di database pola, jika terdapat pola yang sesuai dengan yang ada di database maka ucapan akan dikenali, dan jika tidak terdapat pola yang sesuai dengan yang ada di database maka ANN akan mempelajari pola baru dan disimpan di database sebagai data pola baru, proses belajar dan mengenali ucapan tersebut seterusnya berlangsung pada setiap ada sinyal ucapan yang masuk. Sehingga mesin akan selalu belajar dan mengenali setiap ucapan yang masuk, dan disimpan di database sebagai

METODE PENELITIAN

Metoda penelitian yang akan dibahas pada makalah ini menggunakan metoda LPC. Dengan metoda LPC maka akan dihasilkan parameter-parameter suara manusia sesuai dengan bentuk kata yang diucapkan dari beragam pembicara yang berlainan. Parameter-parameter suara manusia tersebut dapat digunakan untuk sistem pengenal ucapan, pengenal pembicara, text-to-speech recognition, dan aplikasi-aplikasi yang membuat manusia dapat berinteraksi dengan mesin menggunakan sinyal ucapan.

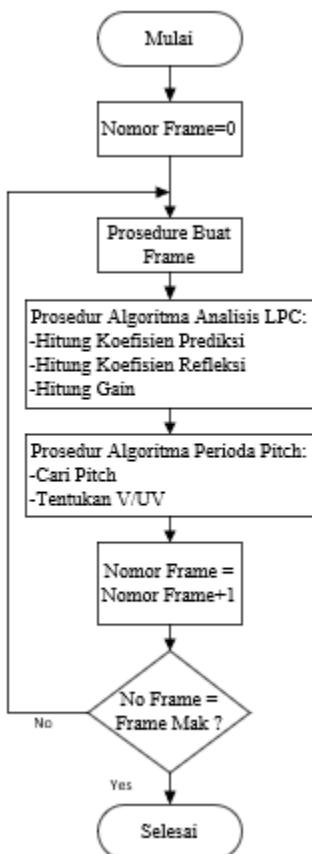
2.1. Ekstraksi sinyal ucapan.

Untuk mengekstraksi sinyal ucapan agar dapat menghasilkan parameter-parameter suara manusia yang unik dalam makalah ini akan menggunakan perhitungan model analisis, ada dua algoritma perhitungan pada bagian analisis, yaitu “algoritma analisis LPC” dan “algoritma perioda

pitch". Algoritma analisis LPC menghasilkan koefisien filter, dan gain. Sedangkan algoritma pitch menghasilkan perioda pitch dan parameter voice/unvoice. Diagram alir pembentukan parameter-parameter LPC ditunjukkan di gambar 4, dengan prosedur yang terdiri dari :

- 1) Prosedur buat frame
- 2) Prosedur algoritma analisis LPC, terdiri dari :
 - a. hitung koefisien prediksi
 - b. hitung koefisien refleksi
 - c. hitung gain
- 3) Prosedur algoritma perioda pitch, terdiri dari :
 - a. cari perioda pitch,
 - b. tentukan Voice/Unvoice.

Dari hasil penelitian (Markel, 1976) untuk sinyal suara manusia, panjangnya frame berkisar antara 15 ms – 40 ms. Misalnya untuk data ucapan sekitar 1 detik dan panjang frame yang digunakan sebesar 30 ms akan terdapat sekitar 33 frame.



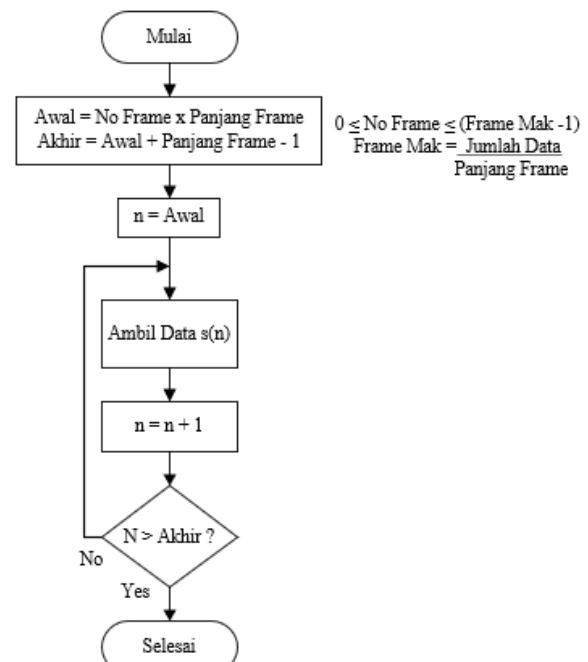
Gambar 4. Prosedur Ekstraksi Sinyal Ucapan

2.1.1. Prosedur buat frame.

Jika frekuensi cuplik yang digunakan adalah sebesar 8 KHz, maka untuk data ucapan sepanjang 1 detik akan terdapat sebanyak 8000 titik data cuplikan, dan jika panjang frame yang digunakan sebesar 30 ms maka akan terdapat sebanyak 240 titik data cuplikan untuk satu frame.

Dengan demikian prosedur buat frame adalah sbb:

Jika data cuplikan pertama diberi nomor $n = 0$, maka untuk frame pertama, data cuplikan diambil dari $n = 0$ sampai $n = (240 - 1)$. Untuk frame kedua, data cuplikan diambil dari $n = 240$ sampai $n = (480 - 1)$. Untuk frame selanjutnya dilakukan dengan cara yang sama dengan diatas sampai dengan frame yang ke 33. Diagram alir prosedur buat frame diperlihatkan di gambar 5.



Gambar 5. Prosedur Buat Frame

2.1.2. Prosedur algoritma analisis LPC.

Prosedur ini adalah untuk menghitung koefisien prediksi, koefisien refleksi, dan gain dengan menggunakan metoda autokorelasi. Diagram alir algoritma autokorelasi dapat dilihat di gambar 6, dengan langkah-langkah prosedur sbb :

1. Hitung autokorelasi data cuplikan sesuai dengan persamaan berikut:

$$r_i = \left(\frac{1}{N} \right) \sum_{n=0}^{N-1-i} s(n)s(n-i) s(n), i = 0, 1, 2, \dots, p.$$

dimana : p adalah jumlah orde prediksi,
 r_i adalah sample autokorelasi ke $-i$,
 N adalah jumlah data cuplikan dalam satu frame,
 $s(n)$ adalah data cuplikan saat ke- n ,
 $s(n-i)$ adalah data cuplikan saat ke- $(n-i)$.

2. Untuk tahap awal $p = 0$, dipilih koefisien prediksi $a_{p=0}(0)=1$, karena itu dari persamaan berikut:

$$E_p = \sum_{i=0}^p a(i).r_i$$

diperoleh: $E_0 = r_0$

3. Untuk tahap p selanjutnya ($1 \leq p \leq$ orde max) :
- dipilih $a_p(0) = 1$,
 - dari persamaan berikut hitung koefisien refleksi :

$$k_p = \frac{-1}{E_{p-1}} \sum_{i=0}^{p-1} a_{p-1}(i) r_{p-i}$$

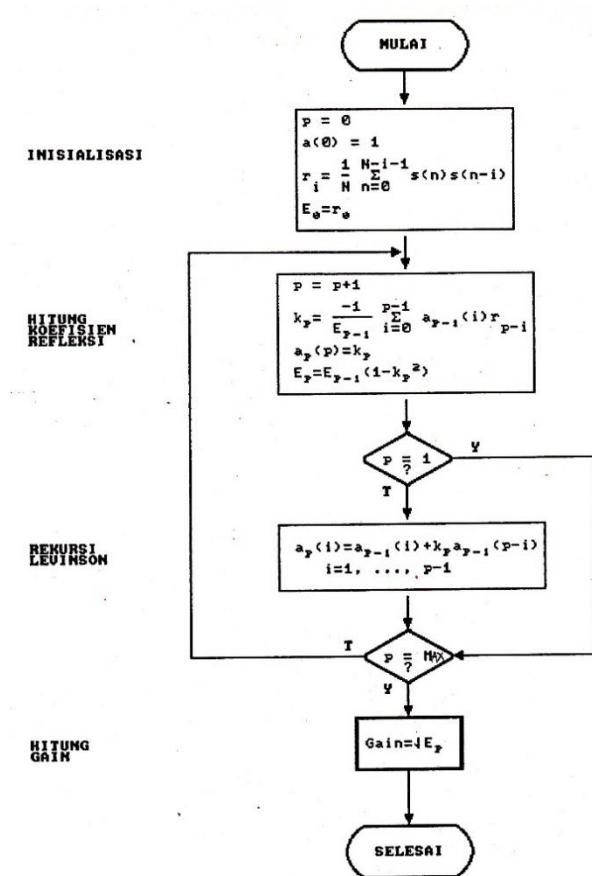
- koefisien prediksi $a_p(p) = k_p$
- Dari persamaan berikut hitung energi error predikasi :

$$E_p = E_{p-1}(1 - k_p^2)$$

4. Koefisien prediksi yang baru dihitung dari persamaan berikut :

$$a_p(i) = a_{p-1}(i) + k_p a_{p-1}(p-i)$$

5. Hitung gain $G = \sqrt{E_p}$



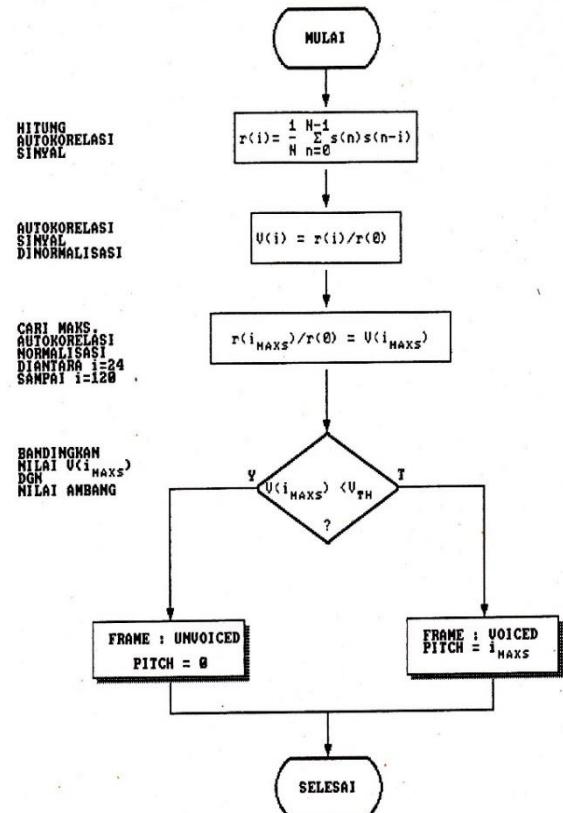
Gambar 6. Prosedur Analisis algoritma LPC

2.1.3. Prosedur algoritma perioda pitch.

Prosedur ini untuk menghitung perioda pitch dan penentuan jenis bunyi Voice/Unvoice. Diagram alir prosedur algoritma perioda pitch dapat

dilihat di gambar 7. dan langkah-langkahnya seperti yang diusulkan oleh (Rabiner L. R., 1977) :

1. Hitung autokorelasi sinyal menurut persamaan :



Gambar 7. Prosedur algoritma Perioda Pitch

$$r_i = \left(\frac{1}{N} \right) \sum_{n=0}^{N-1} s(n)s(n-i), \quad i = 0, 1, 2, \dots, p.$$

2. Harga autokorelasi hasil perhitungan di langkah ke-1 dinormalisasi terhadap harga autokorelasi sinyal pada $i = 0$, $r(i)/r(0)$.

3. Cari harga maksimum di langkah ke-2 pada daerah antara $i = 24$ sampai $i = 120$, yaitu dengan menganggap perioda pitch untuk ucapan manusia sekitar 3 mdetik sampai 15 mdetik dan jika frekuensi cuplik yang digunakan sebesar 8 KHz.

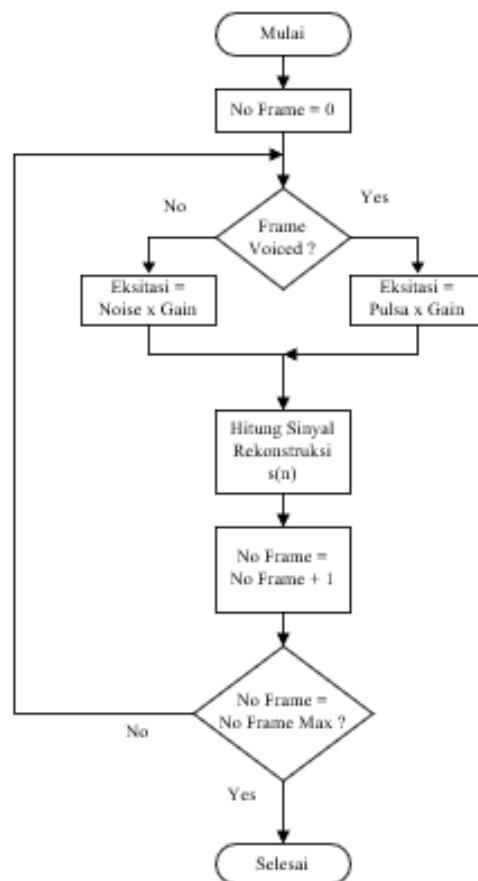
4. Harga maksimum di langkah ke-3 tersebut dibandingkan dengan suatu harga ambang tertentu :

- Jika $r(i_{max})/r(0) < 0,25$, maka frame tersebut diputuskan sebagai jenis bunyi tak-bergetar dengan perioda pitch dibuat sama dengan nol.
- Jika $r(i_{max})/r(0) \geq 0,25$, maka frame tersebut diputuskan sebagai jenis bunyi bergetar dengan perioda pitch sama dengan waktu pada saat harga maksimum terjadi (i_{max}).

2.2. Pengenalan ucapan.

Program ini berfungsi merekonstruksi kembali sinyal ucapan berdasarkan parameter-parameter yang dibaca dari data ekstraksi sinyal ucapan. Prosedur sistesis akan melakukan sintesis pada setiap frame yang dibaca dari data ekstraksi sinyal

ucapan tersebut, mulai dari frame pertama sampai frame terakhir. Jika suatu frame merupakan jenis bunyi bergetar, maka sumber eksitasi yang akan dibangkitkan adalah pembangkit pulsa. Jika suatu frame merupakan jenis bunyi tak bergetar, maka pembangkit desis yang akan dibangkitkan. Keluaran sumber eksitasi tersebut diperkuat oleh pengali gain, dan selanjutnya dibentuk (diartikulasi) oleh filter yang berfungsi sebagai rongga pengubah getaran

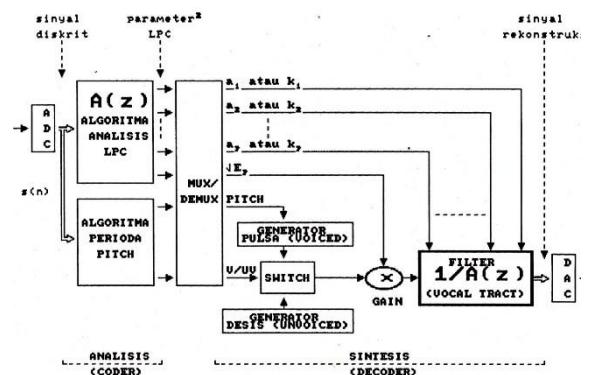


untuk merekonstruksi kembali sinyal ucapan. Diagram alir prosedur pengenal ucapan diperlihatkan di gambar 8.

Gambar 8. Prosedur Pengenal Ucapan

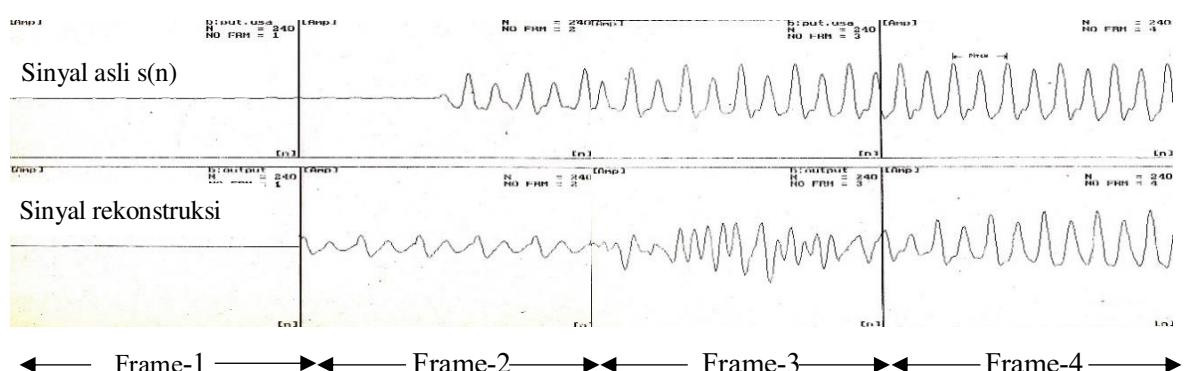
HASIL DAN PEMBAHASAN

Blok diagram sistem analisis dan sintesis LPC diperlihatkan pada gambar 9. Sinyal $s(n)$ adalah sinyal diskrit hasil digitalisasi sinyal kontinyu ucapan, $s(n)$ yang diolah oleh algoritma analisis LPC menjadi parameter-parameter vocal track sebanyak $p = 10$ (koefisien refleksi k_1 s/d k_{10}), dan parameter gain. Sedangkan algoritma perioda pitch akan mengolah $s(n)$ menjadi parameter perioda pitch dan voice/unvoice. Dibagian pengenalan ucapan akan mensintesis parameter-parameter tersebut dengan menggunakan filter vocal tract $1/A(z)$, sehingga dihasilkan sinyal rekonstruksi yang mirip dengan sinyal ucapan diskrit $s(n)$.



Gambar 9. Analisis - Sintesis LPC

Pada gambar 10 terlihat sinyal asli $s(n)$ dibandingkan dengan sinyal rekonstruksi hasil pengenalan ucapan “usai” dari seorang laki-laki berumur 31 tahun pada frame 1, 2, 3, dan 4. Jumlah frame untuk ucapan “usai” ini berjumlah 33 frame tetapi tidak semuanya ditunjukkan di makalah ini, panjang tiap frame 240 sample atau 30 ms dengan frekuensi sampel 8 KHz, dan sinyal rekonstruksi dengan bit-rate 2,4 KBit/det.



Gambar 10. Perbandingan sinyal asli $s(n)$ dengan sinyal rekonstruksi pada frame ke-1, 2, 3, dan 4.

Tabel 1. Konversi ASCII dari sinyal asli s(n) dan sinyal rekonstruksi pada frame ke-1, 2, 3, dan 4.

sinyal asli s(n)	sinyal rekonstruksi
Frame Ke : 1	
Koef. k-1 : -9.9259543850E-01	Koef. k-1 : -9.9215686275E-01
Koef. k-2 : 9.8406229919E-02	Koef. k-2 : 9.5238095238E-02
Koef. k-3 : 1.2098523356E-01	Koef. k-3 : 1.290322806E-01
Koef. k-4 : 7.2398482265E-02	Koef. k-4 : 6.45161290E-02
Koef. k-5 : 4.8741537396E-02	Koef. k-5 : 6.66666666667E-02
Koef. k-6 : -3.66195831115E-02	Koef. k-6 : -6.66666666667E-02
Koef. k-7 : -1.173699217E-01	Koef. k-7 : -1.33333333333E-01
Koef. k-8 : -5.9683411510E-02	Koef. k-8 : -6.66666666667E-02
Koef. k-9 : -2.1336634826E-02	Koef. k-9 : 0.0000000000E+00
Koef. k-10 : -4.3243044710E-02	Koef. k-10 : -6.66666666667E-02
Gain [1] = 2.3791713025E-04	Gain [1] = 0.0000000000E+00
Pitch [1] = 0	Pitch [1] = 0
V/UV [1] = FALSE	V/UV [1] = FALSE
Frame Ke : 2	
Koef. k-1 : -9.3693298045E-01	Koef. k-1 : -9.3725490196E-01
Koef. k-2 : 8.64051631635E-01	Koef. k-2 : 8.5714285714E-01
Koef. k-3 : 1.2667507277E-01	Koef. k-3 : 1.290322806E-01
Koef. k-4 : -1.4453216597E-01	Koef. k-4 : -1.290322806E-01
Koef. k-5 : -9.4037606503E-02	Koef. k-5 : -6.66666666667E-02
Koef. k-6 : 6.4716274950E-02	Koef. k-6 : 6.66666666667E-02
Koef. k-7 : 1.8967044090E-01	Koef. k-7 : 2.0000000000E-01
Koef. k-8 : 2.2754978511E-01	Koef. k-8 : 2.0000000000E-01
Koef. k-9 : 2.0336397886E-01	Koef. k-9 : 2.0000000000E-01
Koef. k-10 : 1.3771699376E-01	Koef. k-10 : 1.33333333333E-01
Gain [2] = 7.9764280184E-03	Gain [2] = 8.0000000000E-03
Pitch [2] = 47	Pitch [2] = 47
V/UV [2] = TRUE	V/UV [2] = TRUE
sinyal asli s(n)	sinyal rekonstruksi
Frame Ke : 3	
Koef. k-1 : -9.4172242777E-01	Koef. k-1 : -9.4117647059E-01
Koef. k-2 : 9.3279450998E-01	Koef. k-2 : 9.3650793651E-01
Koef. k-3 : -2.12503645782E-01	Koef. k-3 : -2.2500623801E-01
Koef. k-4 : -2.0640528712E-01	Koef. k-4 : -1.9354838710E-01
Koef. k-5 : 3.2549013817E-02	Koef. k-5 : 0.0000000000E+00
Koef. k-6 : 2.1444908998E-01	Koef. k-6 : 2.0000000000E-01
Koef. k-7 : 2.9461263945E-01	Koef. k-7 : 2.66666666667E-01
Koef. k-8 : 3.0438762247E-01	Koef. k-8 : 3.33333333333E-01
Koef. k-9 : 2.5839067215E-01	Koef. k-9 : 2.66666666667E-01
Koef. k-10 : 1.8748323506E-01	Koef. k-10 : 2.0000000000E-01
Gain [3] = 9.3134087284E-03	Gain [3] = 9.33333333333E-03
Pitch [3] = 0	Pitch [3] = 0
V/UV [3] = FALSE	V/UV [3] = FALSE
Frame Ke : 4	
Koef. k-1 : -9.3826447519E-01	Koef. k-1 : -9.3725490196E-01
Koef. k-2 : 9.1233584182E-01	Koef. k-2 : 9.0476190476E-01
Koef. k-3 : -1.4533933863E-01	Koef. k-3 : -1.61290322586E-01
Koef. k-4 : -9.1695998404E-02	Koef. k-4 : -9.67741935498E-02
Koef. k-5 : 7.3486579392E-02	Koef. k-5 : 6.66666666667E-02
Koef. k-6 : 1.2588074819E-01	Koef. k-6 : 1.33333333333E-01
Koef. k-7 : 1.4383815582E-01	Koef. k-7 : 1.33333333333E-01
Koef. k-8 : 2.5523464987E-01	Koef. k-8 : 2.66666666667E-01
Koef. k-9 : 3.2821541750E-01	Koef. k-9 : 3.33333333333E-01
Koef. k-10 : 3.2291533350E-01	Koef. k-10 : 3.33333333333E-01
Gain [4] = 1.3264307582E-02	Gain [4] = 1.33333333333E-02
Pitch [4] = 43	Pitch [4] = 43
V/UV [4] = TRUE	V/UV [4] = TRUE

Pada frame ke-4 di gambar 10 terlihat bahwa jumlah perioda pitch ada sekitar 5,5 buah perioda, sehingga besar perioda tersebut kira-kira 240/5,5 sample atau sekitar 43 sampel dan ini sesuai dengan hasil perhitungan frame ke-4 yang diperlihatkan oleh tabel 1. Untuk frame ke-3 terjadi salah interpretasi oleh sistem analisis-sintesis LPC, yang secara visual berjenis voiced akan tetapi hasil perhitungan oleh sistem analisis-sintesis LPC menghasilkan frame berjenis unvoiced seperti yang diperlihatkan di frame ke-3 di tabel 1 oleh parameter v/uv = false yang artinya berjenis unvoiced dengan perioda pitch = 0.

Keluaran pengenal ucapan adalah sinyal rekonstruksi yang besaran parameter-parameternya seperti yang ditunjukkan oleh tabel 1, sinyal rekonstruksi tersebut digambarkan di gambar 3.2. Pada frame ke-2 pada sinyal asli di gambar 10 tampak sebagian berjenis unvoiced dan sebagian lagi berjenis voiced, sinyal diskrit s(n) pada frame ke-2 tersebut dianalisis oleh algoritma perioda pitch menghasilkan parameter v/uv = True yang berarti berjenis voiced dengan perioda pitch = 47 seperti yang ditunjukkan oleh tabel 1. Selanjutnya besaran parameter-parameter pada frame ke-2 tersebut disintesis oleh sistem pengenal ucapan menghasilkan sinyal rekonstruksi yang diinterpretasikan sebagai jenis voiced dengan perioda pitch kira-kira sebesar 47 sample seperti

terlihat di frame ke-2 pada gambar 10. Dengan demikian pada satu frame proses analisis-sintesis hanya dapat menginterpretasikan salah satu dari dua jenis parameter v/uv, jika v/uv = True diinterpretasikan sebagai jenis voiced dan jika v/uv = false diinterpretasikan sebagai jenis unvoiced. Artinya pada proses analisis-sintesis tidak dapat menginterpretasikan kedua jenis parameter v/uv sekaligus dalam satu frame sebagaimana sinyal aslinya, hal ini menyebabkan distorsi.

Dari hasil pengamatan di gambar 10 dan table 1 tampak bahwa sinyal rekonstruksi hasil pengolahan pengenal ucapan tidak persis bentuknya dengan sinyal aslinya, tetapi sinyal rekonstruksi tersebut mengandung informasi pesan ucapan yang sama, sehingga kualitas suaranya pun tentu tidak akan sama persis tetapi masih memiliki kualitas suara yang dapat dimengerti (*intelligible*). Untuk mengurangi distorsi seperti yang terjadi pada frame ke-2, maka dapat diatasi dengan membuat panjang frame yang lebih pendek agar dalam satu frame hanya ada satu jenis v/uv, sehingga salah interpretasi dan salah rekonstruksi dapat diperkecil, namun konsekuensinya akan menambah besar jumlah bit pengkodean parameter-parameternya sehingga meningkatkan bit-rate dan berakibat meningkat juga jumlah bit yang harus diolah dan disimpan dalam memori. Untuk pengukuran kata-kata lainnya seperti ucapan “kata”, “obat”, “ikan”, “ekor”, “enggan”, “biru”, “aneka”, “sate”, “petugas”, “sekolah”, “api”.

KESIMPULAN

Dari hasil pengukuran dan pengamatan tampak bahwa metoda LPC dapat mengekstrasi sinyal ucapan menjadi parameter-parameter yang unik untuk tiap ragam ucapan dan tiap ragam orang sehingga dengan ragam pola parameter-parameter tersebut dapat digunakan untuk membuat aplikasi pengenal suara (speech recognition), aplikasi pengenal pembicara (speaker recognition), aplikasi text-to-speech, dan aplikasi-aplikasi yang dapat membuat mesin dapat berinteraksi dengan manusia yang salah satu cara belajar mengenalinya menggunakan jaringan syaraf tiruan ANN yang merupakan salah satu cabang ilmu kecerdasan buatan AI.

REFERENSI

- Jurafsky, D. &. (2018). *Speech and Language Processing*. Third Edition draft.
- Kala, A. &. (2015). Speech Analysis and Synthesis using Vocoder. *International Journal For Trends In Engineering & Technology*.
- Markel, J. a. (1976). *Linear Prediction of Speech*. New York: Springer - Verlag.
- Muslich, M. D. (1990). *Garis-garis Besar Tatabahasa Baku Bahasa Indonesia*. Malang: YA3 Malang.
- Rabiner, L. R. (1977). On the Use of Autocorrelation Analysis for Pitch Detection. *IEEE Trans. On Acoustics, Speech, and Signal Processing*, Vol. ASSP-25, No.1 Feb.
- Rabiner, L. R. (2007). *Introduction to Digital Speech Processing*. Foundations and Trends in Signal Processing.
- Siang., J. J. (2005). *Jaringan syaraf tiruan dan pemrogramannya menggunakan matlab*. Yogyakarta: Andi.