

BAB II

LANDASAN TEORI

2.1. PROSES PEMBENTUKAN SINYAL WICARA

Sama seperti wajah, jemari atau sosok orang secara keseluruhan, suarapun mempunyai identitas. Lewat suara dapat diketahui usia pembicara, pendidikan dan gaya hidupnya, bahkan dapat dipakai sebagai identifikasi seseorang. Untuk memahami bagaimana hal ini bisa terbentuk, kita perlu mengetahui bagaimana suara manusia diproduksi.

Waktu berbicara, manusia menggunakan Diafragma (Sekat rongga badan antara dada dan perut) untuk mengalirkan udara ke Paru-paru. Kemudian udara ini akan naik melalui saluran Tenggorokan, lalu mengenai Pita suara sehingga Pita suara ini akan bergetar dan menghasilkan suara. Pita suara sendiri adalah jaringan otot mirip gendang yang robek separuh pada bagian tengahnya. Getaran yang diproduksi oleh Pita suara akan menuju dua arah. Yang pertama akan mengalir melalui lubang tenggorokan kemudian keluar lewat lubang hidung, sedang yang kedua mengalir keluar dari Tenggorokan lewat mulut. Didalam mulut, Gigi dan Lidah menjadi filter sekaligus sebagai resonator.

2.2. MEKANISME PEMBENTUKAN SINYAL WICARA

Alat ucap manusia terdiri dari (J. D. Markel and A. H. Gray, 1976) :

a. Pita suara (*vokal cord*)

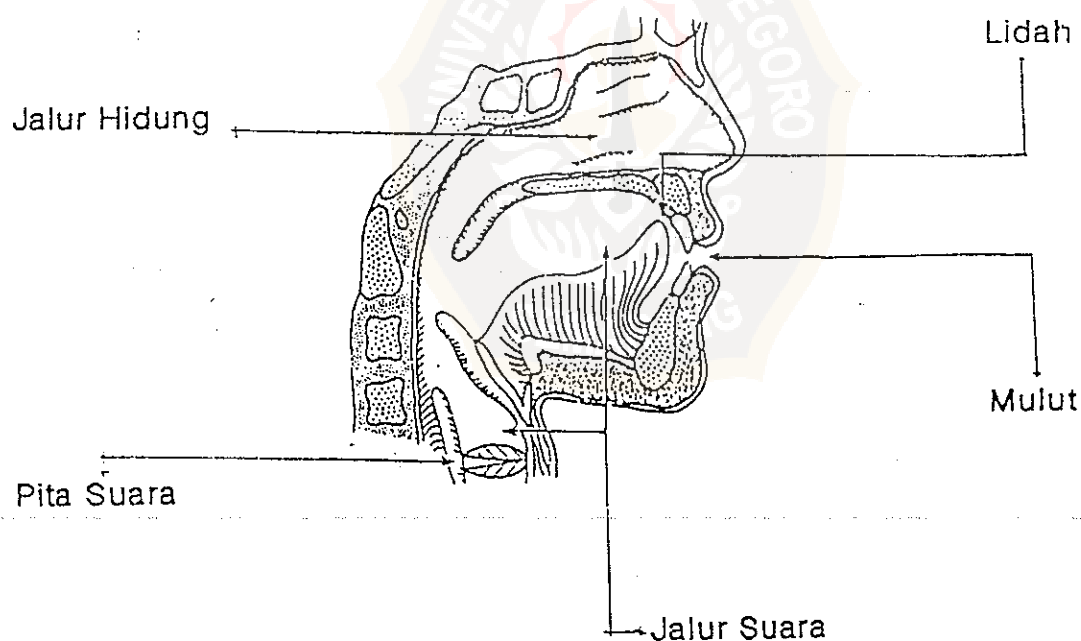
Pita suara terletak melintang pada batang tenggorokan dengan celah diantara (glotis) yang dapat diatur untuk bergetar atau tidak apabila dilalui hembusan udara.

b. Jalur suara (*vokal track*)

Jalur suara diawali dari pita suara, rongga mulut dengan organ-organnya (seperti gigi dan lidah) dan berakhir di bibir.

c. Jalur hidung

Jalur hidung diawali dari katup penutup (velum) rongga hidung dan berakhir dilubang hidung.



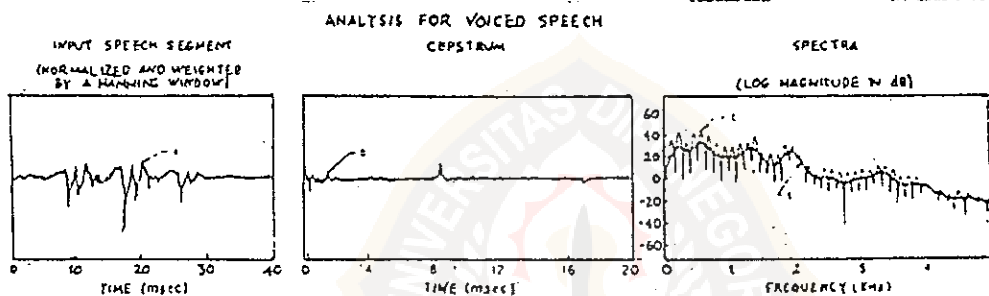
penampang lintang alat wicara manusia

gambar 2.1

Berdasarkan cara suara dihasilkan, maka suara wicara manusia dapat dibedakan menjadi 3 bagian yaitu :

1. Bunyi bersuara (*Voiced*)

Bunyi ini ditimbulkan akibat eksitasi jalur suara oleh sinyal-sinyal aquasi periodik. Sinyal aquasi periodik ini dihasilkan oleh getaran pita suara dimana periodenya sering disebut periode pitch. Contohnya bunyi /u/, /w/, /i/, /e/, /o/, /a/, dsb.

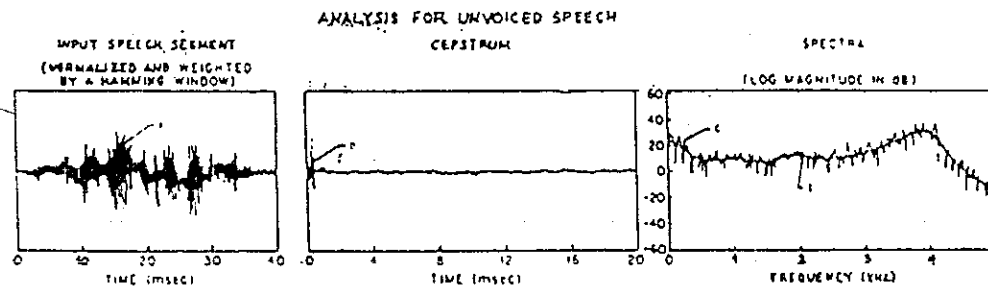


gambar 2.2a

2. Bunyi tak bersuara (*Unvoiced sound*)

Pita suara tak bergetar , bunyi ini dibangkitkan oleh adanya suatu penyempitan pada beberapa titik tertentu dalam jalur suara. Tekanan udara yang melalui lubang penyempitan mempunyai kecepatan yang cukup tinggi untuk menghasilkan turbelensi. Proses ini akan menghasilkan noise yang mengeksitasi jalur suara. Turbelensi ini bersifat acak dan jalur suara-lah yang mengatur turbulensi tadi, artinya untuk menghasilkan huruf 'f' misalnya, jalur suara harus memiliki bentuk yang

berbeda dengan bentuknya yang dihasilkan oleh bunyi 's'. Bunyi tak bersuara (*unvoiced sound*) yang seperti desis / derau acak seperti bunyi /s/ atau 'sh'.



gambar 2.2b

3. Bunyi Letupan

Bunyi ini dihasilkan akibat letupan dari aliran udara yang dihambat sementara waktu pada satu tempat sehingga tekanan (P) naik, kemudian dilepaskan secara tiba-tiba, contohnya bunyi /ts/ atau 'ch'.

Dalam bahasa, satuan bunyi yang masih mampu membedakan arti disebut *Fonem*, dengan hanya memperhatikan bentuk sinyalnya terutama bunyi-bunyi bersuara (*Voiced sound*). Untuk fonem yang sama, bentuknya selalu konstan dan berulang sebesar periode *Pitch* yaitu periode getar pita suara. Dari gambar 2.2.a dapat kita lihat bahwa sinyal fonem bersuara (vokal) berbentuk quasi periodik. Sinyal untuk konstanta dan bunyi desah berbentuk noise (random). Juga *Voiced Speech* mempunyai Amplitudo yang jauh

lebih besar dibanding Amplitudo *Unvoiced Speech*.

Faktor-faktor yang membedakan warna suara manusia satu dengan yang lain adalah posisi Forman dan Periode Pitch. Jalur suara dapat digambar sebagai pipa organa pembuka dengan panjang normal (saat tidak aktif) kurang lebih 17 cm. Bila kita mengucapkan satu kata, panjang jalur suara otomatis berubah sesuai dengan fonem yang diucapkan. Akibatnya frekuensi resonansi sinyal berubah-ubah. Frekuensi-frekuensi resonansi jalur suara ini lazim disebut sebagai *Forman*. Dan tiga Forman yang pertama adalah yang sangat penting, dan diberi simbol F_1 , F_2 , F_3 dan F_0 (frekuensi dasar). Jadi yang menentukan bunyi ujaran adalah posisi Forman dan Periode Pitch.

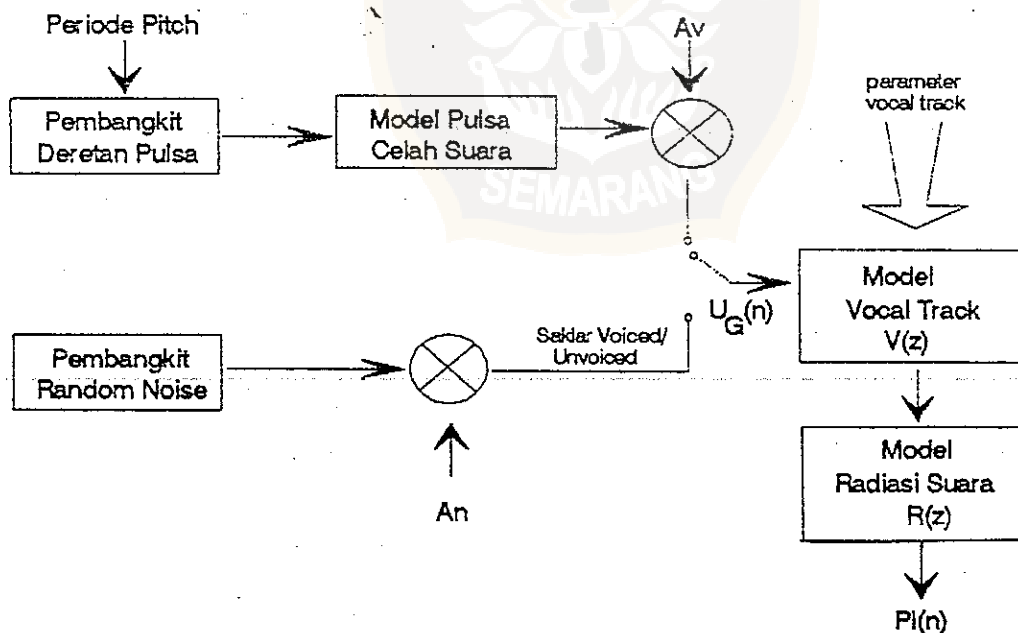
Dari penyelidikan dengan foto sinar-X terlihat bahwa alat ucap manusia dapat dianggap tetap bentuknya selama mengucapkan fonem yang sama. Jadi dalam suatu wicara yang mengandung banyak macam fonem, bentuk alat ucap manusia dianggap konstan untuk waktu yang pendek, yaitu selama periode pitch. Dari hasil inilah maka Penulis menggunakan metode LPC (*Linear Predictive Coding*) dimana analisa sinyal dilakukan dalam selang yang pendek (*short time analisis*). Ada beberapa hal yang mempengaruhi pembentuk suara (J. D. Markel and A. H. Gray, 1976) , yaitu:

1. Perubahan bentuk jalur suara terhadap waktu.
2. Rugi-rugi energi sebagai akibat konduksi panas dan gesekan pada jalur suara.
3. Elastisitas dinding jalur suara.

4. Radiasi suara pada bibir.
5. Hubungan dengan jalur hidung.
6. Jenis eksitasi yang menyebabkan bunyi.

2.3. MODEL DIGITAL SINYAL WICARA

Pengaruh dari hal-hal yang telah disebutkan diatas terhadap wicara yang dihasilkan dapat dimodelkan menjadi filter digital yang berubah terhadap waktu. Ada 3 aspek utama, yaitu radiasi udara pada bibir dengan fungsi pindahanya $R(z)$, jalur suara dengan fungsi pindahanya $V(z)$ dan eksitasi pada pita suara yang mempunyai fungsi pindah $G(z)$. Sedangkan keras lemahnya suara dapat diatur melalui parameter A (gain). Gain untuk wicara *Voiced* (A_v) maupun wicara *Unvoiced* (A_u).



model digital sinyal wicara

gambar 2.3.

Dalam analisis sinyal wicara diperlukan model matematis dari alat ucap manusia, yaitu jalur suara. Model yang dipilih harus cukup sederhana tetapi mencerminkan mekanisme pembentukan sinyal wicara yang teliti, maka dengan mengetahui parameter-parameter model, dapat diprediksi bunyi yang timbul.

Pada saat melakukan proses berbicara, bentuk jalur suara berubah secara kontinyu, demikian pula kondisi sumber eksitasi berubah dari waktu ke waktu. Bentuk jalur suara serta kondisi sumber eksitasi diasumsikan tetap konstan selama selang waktu yang sangat pendek, kira-kira 10 - 20 ms (L.R. Rabiner and R.W. Schafer, 1978).

2.3.1. Jalur Suara (Vokal Track)

Jalur suara dan organ artikulasinya dimodelkan oleh persamaan All-pole ($V(z)$) yang parameternya quasi stasionernya (parameter $V(z)$) diperbaharui tiap 10-20 ms sekali (L.R. Rabiner and R.W. Schafer 1978). Persamaan $V(z)$ memenuhi bentuk :

$$V_z = \frac{G}{\sum_{k=0}^p a_k z^{-k}}$$

dengan :

G = penguat

a_k , $0 \leq k \leq p$ = koefisien prediksi

z^{-k} $0 \leq k \leq p$ = delay ke-k

p = orde prediksi

2.3.2 RADIASI

Radiasi udara pada bibir merupakan salah satu aspek yang mempengaruhi bentuk model digital sinyal wicara. Efek radiasi ini merupakan efek rugi-rugi karena radiasi suara pada bibir (L.R. Rabiner and R.W. Schafer 1978). Pada model digital sinyal wicara, radiasi wicara pada bibir dilambangkan dengan fungsi pindah $R(z)$

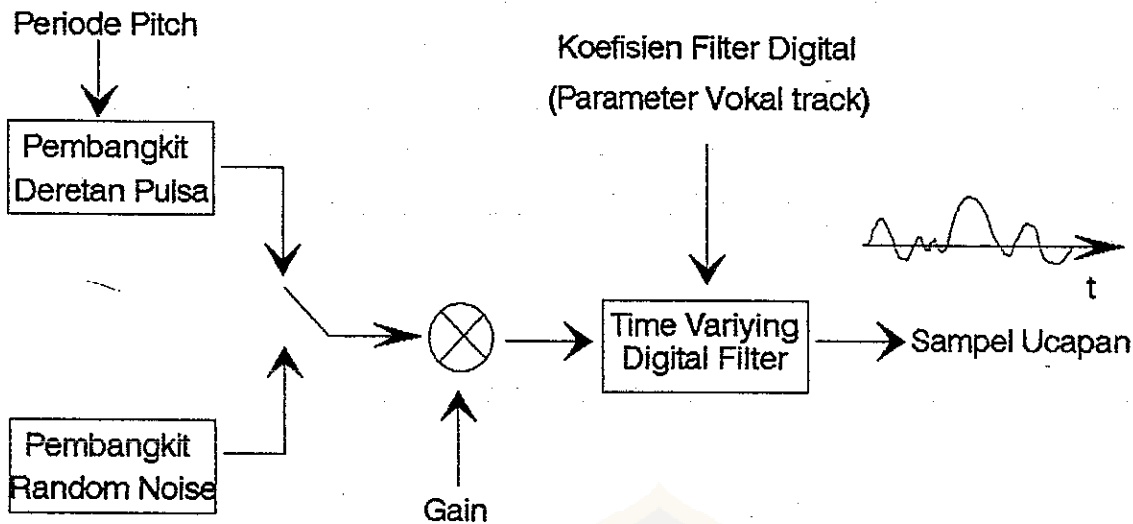
2.3.3. Eksitasi

Sumber eksitasi untuk model ini ada 2 macam, yaitu sumber eksitasi berupa pulsa-pulsa periode dan sumber eksitasi berupa derau acak. Contoh bentuk gelombang *voiced sound* dan *unvoiced sound* dapat dilihat pada gambar 2.2a. dan gambar 2.2b., terlihat untuk ucapan *voiced* bentuknya periodik dengan periode sama, sedangkan untuk ucapan *unvoiced* bentuknya derau acak.

2.4. PENDEKATAN TERHADAP MODEL DIGITAL

Pada aplikasi menggunakan metode LPC, model seperti gambar 2.3. dapat disederhanakan lebih lanjut dengan menggabungkan pengaruh radiasi pada bibir, jalur suara dan eksitasi pada pita suara menjadi filter digital tunggal. Filter ini juga berubah terhadap waktu dan fungsi transfernya hanya mengandung pole yang konstan selama selang yang singkat. (lihat bagan pendekatan terhadap Model Digital) (T. W. Parson, 1982)

BAGAN PENDEKATAN MODEL DIGITAL



1. Pembangkit Deretan Pulsa

Pembangkit ini adalah pembangkit deretan pulsa dengan periode ditentukan oleh periode pitch sebagai sumber eksitasi untuk wicara *voiced*.

2. Pembangkit suara acak (*Random Noise*)

Pembangkit ini adalah pembangkit noise acak sebagai sumber eksitasi untuk wicara *unvoiced* (L. R. Rabiner and R. W. Schafer, 1978).

3. Saklar *voiced / unvoiced*

Saklar ini adalah penentu jenis eksitasi apabila saklar ini pada posisi ke atas maka sumber eksitasi adalah pembangkit deretan pulsa, sedangkan pada posisi saklar kebawah maka sumber eksitasi adalah pembangkit suara acak.

4. Gain

Penentu keras lemahnya suara.

5. Time Varying Digital Filter

Frekuensi resonansi rongga mulut dan hidung dapat diatur melalui parameter tertentu agar diperoleh wicara yang diinginkan pada keluarannya.

Dari gambar 2.4. dapat diketahui parameter-parameter yang mempengaruhi karakteristik wicara, yaitu :

1. Penguat G

Mempengaruhi kuat lemahnya suara

2. Periode Pitch

Menentukan tinggi rendahnya nada suatu wicara. Secara umum menyebabkan suara wanita terdengar melengking dan sebaliknya suara pria terdengar lebih berat

3. Klasifikasi *Voiced dan Unvoiced*

Klasifikasi ini diperlukan untuk menentukan suatu wicara termasuk wicara *voiced* atau wicara *unvoiced*.

4. Koefisien Prediksi

Parameter ini menentukan wicara yang dihasilkan, terutama sangat berguna untuk bunyi vokal

Keadaan quasi stasioner yang dimiliki oleh sinyal wicara sangat menguntungkan bila dikaitkan dengan kebutuhan memori karena hal ini berarti penghematan jumlah data yang harus disimpan. Meskipun begitu, secara ideal parameter-parameter $V(z)$ haruslah merupakan variabel yang tergantung waktu. Ini berarti bahwa salah satu variabel $V(z)$ harus berupa $t(\text{time})$. Ini merupakan salah satu

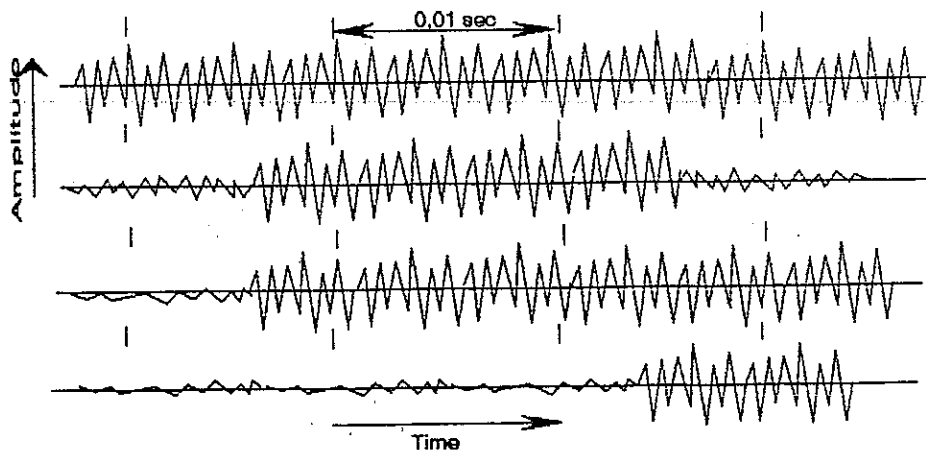
kelemahan dari model yang dipilih.

Untuk wicara *voiced*, model ini sudah bekerja dengan baik, tetapi untuk *unvoiced* terutama suara transien, model ini tidak begitu baik ketelitiannya. Untuk mengatasi hal tersebut, kita harus menggunakan sistem yang kompleks, prosesor yang cepat dan sistem perancangan yang teliti.

(L. R. Rabiner and R. W. Schafer 1978).

2.5. WINDOWING

Suatu sinyal dapat dinyatakan dalam bentuk gelombang, dan hubungan antara gelombang dengan spektrumnya dapat dinyatakan dalam transformasi Fourier. Tetapi transformasi Fourier terlalu sulit untuk perhitungan wicara karena diperlukan data yang cukup besar, sehingga perhitungannya tidak efisien dan batas integrasinya panjang sekali, padahal pada setiap fonem perubahan sinyalnya cepat sekali ($\pm 1 - 10$ ms), maka bisa-bisa informasi-informasi yang penting malah jadi hilang. (lihat contoh bentuk gelombang sinyal wicara).



gambar 2.4.

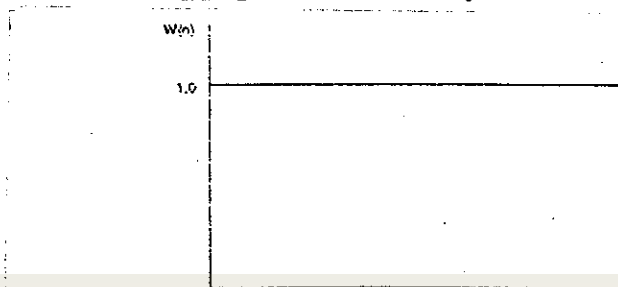
Terlihat bahwa sinyal wicara bersifat stasioner dan seragam untuk interval waktu yang pendek ($\pm 1 - 10$ ms), tetapi berubah drastis dalam selang 0,5 sekon. Kondisi inilah yang disebut *Quasi Stasioner* (S. Saito dan K. Nakata, 1985). Agar kondisi stasionernya terpenuhi, maka sinyal wicara harus di window. Dengan demikian sinyal masukan dibagi-bagi menjadi frame-frame dan tiap frame terdiri n sampel. Frame harus pendek agar kondisi stasionernya terjaga tetapi juga harus cukup panjang agar periode Pitchnya tercakup. Berdasar percobaan-percobaan yang dilakukan didapat selang yang tepat adalah 10-30 ms (S. Saito dan K. Nakata, 1985). Selang tersebut lazim disebut 1 frame. Window yang sering digunakan dalam analisa sinyal wicara adalah *Rectangular Window* dan *Hanning Window*.

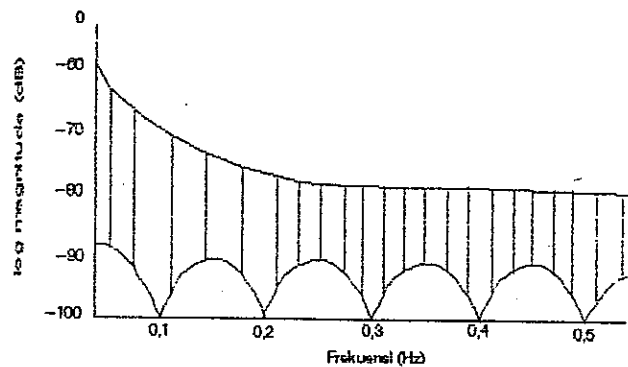
2.5.1. Rectangular Window

Window yang paling sederhana, berbentuk segi empat dengan persamaan :

$$W(n) \begin{cases} = 1 ; & 0 \leq n \leq (N - 1) \\ = a ; & \text{untuk } n \text{ lainnya} \end{cases} \quad [2.2]$$

Pada gambar 2.5. dapat dilihat bentuk fungsi rektanguler window beserta tanggap frekuensinya.



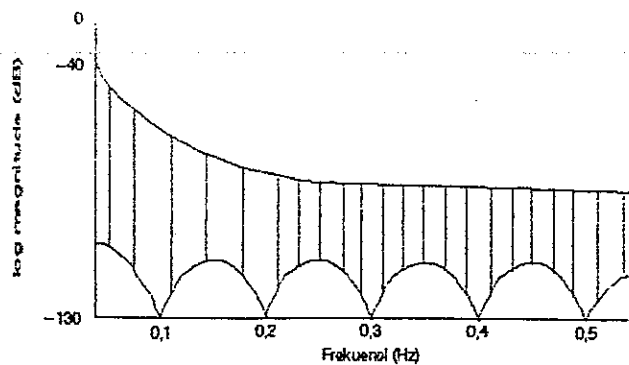
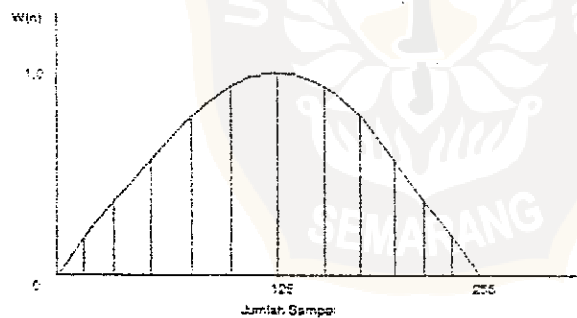


gambar 2.5.

2.5.2. Hamming Window

Hamming window memiliki persamaan (S. Saito dan K. Nakata, 1985) :

$$W(n) \begin{cases} = 0,54 - 0,46 \cos(2\pi n / (N-1)); & 0 \leq n \leq (N-1) \\ = 0; & \text{untuk } n \text{ lainnya} \end{cases} \quad [2.3]$$



gambar 2.6.

Dari kedua fungsi window tersebut, tampak bahwa Rectangular Window memiliki lobe yang lebih sempit dibanding main lobe dari *Hamming Window*. Pada *Hamming Window*, lebar main lobenya dua kali dari main lobe *Rectangular Window*, tetapi *Hamming Window* memiliki side lobe yang lebih kecil terhadap main lobenya dibanding dengan side lobe *Rectangular Window*. Pada skala log, side lobe dari *Rectangular Window* besarnya sekitar -20 db, sedangkan side lobe *Hamming Window* kurang dari -20 db.

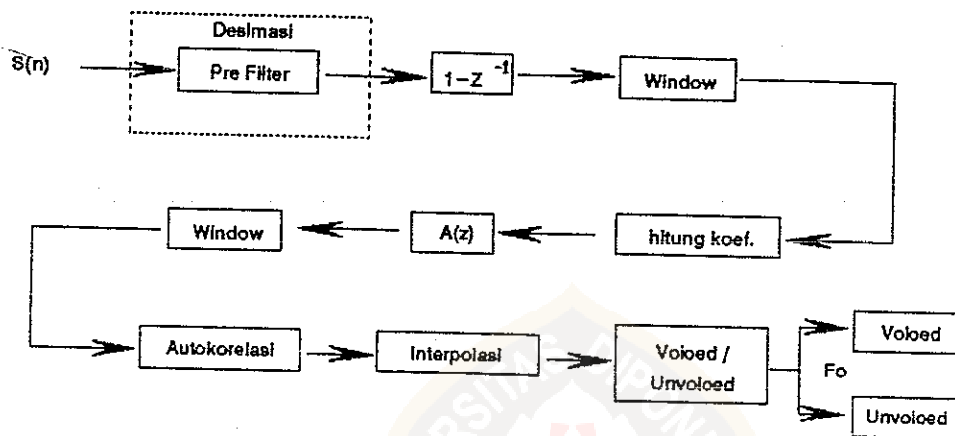
Ditinjau dari sisi jangkauan frekuensinya suatu fungsi Window yang baik yaitu bila memiliki resolusi frekuensi yang tinggi (main lobenya sempit dan curam) dan side lobenya sangat kecil dibanding dengan main lobenya. Untuk idealnya, main lobe suatu fungsi Window haruslah sempit dan curam sekali dengan side lobe sama dengan nol. Fungsi Window yang baik diatas memang merupakan dua hal yang saling bertentangan sehingga tidak akan dapat dipenuhi kedua-duanya sekaligus dengan baik. (S. Saito and K. Nakata, 1985)

2.6. ALGORITMA SIFT

2.6.1. PENGANTAR

Algoritma SIFT (*Simplified Invers Filter Tracking*) merupakan metode yang teliti dan efisien untuk menentukan periode Pitch sinyal bicara pada jangkauan frekuensi 50-250 kHz, berdasar pada perkiraan linier. (J. D. Markel and A. H. Gray)

Pada algoritma ini dilakukan penurunan kecepatan sampling sehingga frekuensi sampling efektif untuk analisa Frekuensi dasar F_0 , yaitu 2 kHz. Ini berarti hanya jangkauan frekuensi sekitar 1 kHz saja yang diproses, sehingga memperkecil jumlah perhitungan.



gambar 2.7.

Pada gambar tersebut diatas, $S(n)$ adalah deretan sample yang diperoleh dari hasil pencuplikan sebesar 8 kHz. Mula-mula sample $S(n)$ diumpankan ke Low Pass Filter yang memiliki frekuensi cut off sebesar 0,8 kHz. Setelah difilter, kemudian dicuplik dengan kecepatan pencuplikan $f_s/I = 2$ kHz, dimana f_s adalah frekuensi cuplikan semula dan I adalah bilangan bulat yang menyatakan faktor down sampling. Proses pre filtering dan down sampling ini disebut *Desimasi*. Kemudian mencari selisih masing-masing sample dengan sample sebelumnya untuk menonjolkan daerah Forman kedua dan dikalikan dengan fungsi *Hamming Window*. Setelah itu dirancang Invers Filter orde-4, dengan empat koefisien sudah cukup memadai, karena sudah mencakup 2

Forman pada jangkauan frekuensi (0-1) kHz, sedang Koefisien-koefisien prediksi yang dapat dicari melalui metode LPC. Setelah koefisien prediksi diketahui, maka keluaran Inverse filter ini dikalikan dengan fungsi *Hamming Window* dan di Autokorelasikan sehingga dihasilkan deretan nilai Autokorelasi dimana Frekuensi dasar F_0 diestimasi. F_0 diestimasi dengan mencari nilai puncak, yaitu nilai Autokorelasi tertinggi pada frame tertentu. Agar didapat hasil yang teliti maka metode metode interpolasi digunakan untuk menghitung puncak tertinggi. Akhirnya penentuan voiced / unvoiced dilakukan berdasar nilai puncak yang diinterpolasi tersebut. Jika framenya voiced, maka letak nilai puncak yang diinterpolasi tersebut menyatakan periode pitch p , sedang jika sebaliknya menyatakan nilai frekuensi dasar F_0 . Bila tidak ditemukan periode pitch p , berarti framenya unvoiced.

2.6.2. DESIMASI

Pada pengolahan sinyal seringkali dilakukan penurunan kecepatan pencuplikan, hal ini dilakukan untuk menghemat memori penyimpanan data hasil pencuplikan dan untuk memperkecil jumlah operasi perhitungan yang dilakukan. Penurunan kecepatan pencuplikan ini disebut *Desimasi*. Misal diinginkan penurunan kecepatan sampling sebesar faktor m dari deret sampel $S(n)$. Untuk itu dibuat deret baru $W(n)$ dengan mengambil secara periodik satu sampel dari setiap m sampel. Jika syarat kriteria Nyquist

dipenuhi , yaitu $1/T > 2 F_N$, maka sampel dari deret baru $W(n)$ dapat digunakan untuk membentuk kembali sinyal analog aslinya.



gambar 2.8

Dari diagram terlihat sinyal $S(n)$ diumpankan ke LPF dengan frekuensi cut off sebesar $\pi/\tau = \pi/3\tau$ (kecepatan sampling diturunkan dengan faktor 3) sehingga menghasilkan deret baru. Setelah pemfilteran, aliasing tidak terjadi meski kecepatan pencuplikan diturunkan dengan faktor 3. Jika syarat kriteria Nyquist dipenuhi, maka pemfilteran awal/prefiltering perlu dilakukan.

2.6.3. Prefiltering

Pada diagram kotak Algoritma SIFT, sinyal masukan $S(n)$ diumpankan lebih dahulu di LPF. Hal ini dilakukan untuk mencegah timbulnya aliasing frekuensi penggal. LPF dipilih sedemikian rupa sehingga syarat Nyquist dipenuhi pada Algoritma SIFT ini, filter yang digunakan adalah digital Chebyser 3 kutub dengan ripple 2 db yang memiliki persamaan (J. D. Markel, 1972) :

$$U(n) = A_1 S(n) + A_2 u(N-1) \quad [2.4]$$

$$X(n) = A_3 U(n) + A_4 X(n-1) + A_5 x(n-2)$$

dengan $A_1 = 1 - \exp(-\alpha_1 T)$

$$A_2 = \exp(-\alpha_1 T)$$

$$A_3 = 1 - 2 \exp(-\alpha z T) \cos \beta z T + \exp(-2\alpha z T)$$

$$A_4 = 2 \exp(-\alpha z T) \cos \beta z T$$

$$A_5 = -\exp(-2\alpha z T)$$

$$\alpha_1 = (0,3572) 2\pi f_c$$

$$\alpha_2 = (0,1786) \pi f_c$$

$$\beta z = (0,8938) \pi f_c$$

$$U(n) = 0, \quad n < 0$$

$$X(n) = 0, \quad n < 0$$

{ S(n) } adalah deretan Input dan { X(n) } adalah deretan Output LFF. Frekuensi penggalnya dinyatakan dengan f_c , sedang T menyatakan periode pencuplikan. Setelah dilewatkan LFF, kemudian dilakukan penurunan kecepatan pencuplikan.

2.7. Filter Invers Digital

Pada sintesa sinyal wicara digunakan model digital pembentuk sinyal wicara yang sudah disederhanakan seperti yang ditunjukkan pada gambar 2.6. Pada model tersebut fungsi pemindah time varying digital filter yaitu $H(z)$ dinyatakan dengan persamaan :

$$H(z) = \frac{G}{1 + \sum_{i=1}^p a_i z^{-i}} \quad [2.5]$$

untuk $G = 1$ maka persamaan tersebut dapat ditulis menjadi

$$H(z) = \frac{1}{1 + \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad [2.6]$$

kebalikan dari filter ini disebut Invers Filter yang memiliki fungsi periode sebagai berikut :

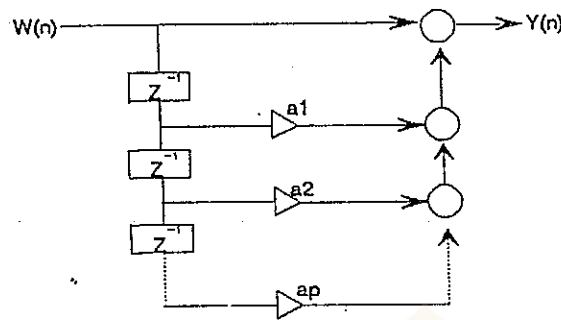
$$H(z)^{-1} = A(z)$$

dengan

$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i}$$

$$= \sum_{i=0}^p a_i z^{-i}$$

[2.7]



gambar 2.9.

Struktur Invers Filter Digital

Persamaan diatas memiliki struktur digital seperti yang ditunjukkan pada gambar 2.9. dengan

a_i , $0 \leq i \leq p$ = koefisien prediksi

z^{-1} , $0 \leq i \leq p$ = delay ke-i

$W(n)$ = menyatakan deretan sampel setelah didesimasi

$Y(n)$ = menyatakan deretan sampel keluar -an Invers Filter

Berdasar persamaan diatas maka deretan sampel keluaran Invers Filter dapat dinyatakan dengan :

$$y(n) = w(n) + \sum_{i=1}^p a_i w(n-i) \quad , \quad n = 0, 1, \dots, (r-1) \quad [2.8]$$

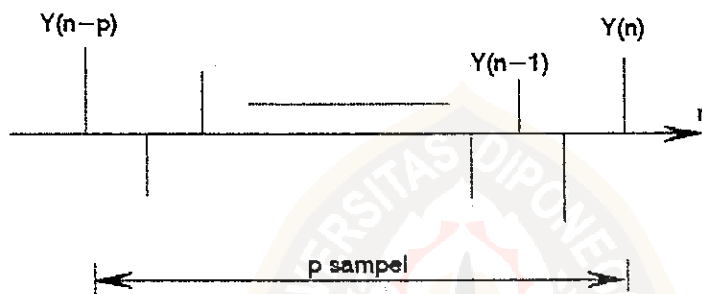
N adalah jumlah sampel dalam satu frame.

Koefesien prediksi a_1 dapat dicari melalui metode Prediksi Linier yang akan dibahas pada bagian berikutnya.

2.8. Prediksi Linier Sinyal Wicara

Suatu sampel sinyal wicara dapat diprediksi dengan kombinasi linier dari sampel-sampel sebelumnya. Pada gambar 2.10. akan diestimasi besarnya sampel $y(n)$ sinyal prediksi $\hat{y}(n)$ dimana $\hat{y}(n)$ merupakan kombinasi linier dari sejumlah p sampel sebelumnya. Sinyal prediksi $\hat{y}(n)$ dinyatakan dengan persamaan sebagai berikut :

$$\hat{y}(n) = - \sum_{i=1}^p a(i) y(n-i) \quad [2.9]$$



Gambar 2.10.

Estimasi Sampel $y(n)$ dengan Kombinasi Linier sebelumnya dengan demikian kesalahan prediksinya dapat dinyatakan dengan :

$$\begin{aligned} e(n) &= y(n) - \hat{y}(n) \\ &= \sum_{i=0}^p a(i) y(n-i) \end{aligned} \quad [2.10]$$

dengan $a(0) = 1$ (untuk memudahkan perhitungan).

Agar diperoleh hasil prediksi yang maksimal, maka error harus minimum. Error prediksi minimum diperoleh dengan menerapkan kriteria least square, yaitu minimum dari kuadrat error.

$$E = \langle e^2 \rangle = \text{minimum}$$

Koefisien $a(i)$ diperoleh dengan membuat agar error prediksi sekecil mungkin. Koefisien-koefisien ini dapat pula dicari dengan cara menerapkan penurunan parsial dari $\langle e^2 \rangle$ terhadap $a(i)$, tetapi akan lebih mudah dan cepat jika digunakan prinsip orthogonalitas. Prinsip ini menyatakan bahwa koefisien-koefisien tersebut adalah koefisien yang membuat errornya orthogonal terhadap sampel $y(n-1), y(n-2) \dots y(n-p)$ (dua fungsi u dan v akan orthogonal jika $\langle u \ v \rangle = 0$). Dengan demikian akan diperoleh :

$$\langle y(n-j) \ e(n) \rangle = 0, \text{ untuk } j = 1, 2, \dots, p \quad [2.11]$$

sehingga dapat ditulis

$$\langle y(n-j) \sum_{i=0}^p a(i) y(n-i) \rangle = 0 \quad [2.12]$$

Dengan menukarkan operasi perata-rataan dan jumlahan serta menyatakan $\langle \ \rangle$ dengan jumlah n maka akan diperoleh

$$\sum_{i=0}^p a(i) \sum_n y(n-i) \langle y(n-j) \rangle = 0, \text{ } j = 1, 2, \dots, p \quad [2.13]$$

Prediktor yang dikehendaki dapat ditentukan dengan menyelesaikan persamaan-persamaan ini. Prinsip orthogonalitas juga menyatakan bahwa error minimum diberikan oleh :

$$\langle e^2(n) \rangle = \langle y(n) - e(n) \rangle \quad [2.14]$$

Dengan cara yang sama seperti cara untuk memperoleh persamaan 2.13. maka dapat ditunjukkan bahwa :

$$\sum_{i=0}^p a(i) \sum_n y(n-i) \langle y(n-j) \rangle = E \quad [2.15]$$

Batas n pada penjumlahan ini tergantung pada cara memindahkan errornya, ada dua cara untuk memindahkannya :
cara pertama :

Error dibuat minimum untuk n titik sinyal wicara. Pada kasus ini batas n adalah 0 , dan $N - 1$, sehingga persamaannya menjadi :

$$\sum_{i=0}^P a(i) c_{ij} = 0, \quad j = 1, 2, 3, \dots, p \quad [2.16]$$

dan
$$E = \sum_{i=0}^P a(i) c_{ij}$$

dengan

$$C_{ij} = \sum_{n=0} y(n-i) y(n-j)$$

Persamaan ini disebut Metode Kovarian.

Untuk menghitung C_{00} sampai C_{pp} dibutuhkan $(N+P)$ sampel sinyal wicara.

cara kedua :

Error dibuat dalam seluruh waktu, dalam hal ini batas n adalah antara $-\infty$ dan ∞ . dengan demikian dapat dituliskan sebagai berikut :

$$\sum_{i=0}^P a(i) r_{ij} = 0, \quad j = 1, 2, 3, \dots, p \quad [2.17]$$

dan
$$E = \sum_{i=0}^P a(i) r_i$$

dengan

$$r_i = \sum_{n=-\infty}^{\infty} y(n) y(n-i)$$

Karena hanya diketahui sejumlah N titik tertentu maka diasumsikan bahwa $y = 0$ diluar daerah $(0, N-1)$ maka diperoleh :

$$r_i = \sum_{n=i}^{n-1} y(n) y(n-i) \quad [2.18]$$

Bentuk persamaan diatas disebut metode autokorelasi, dan fungsi r disebut fungsi autokorelasi dari $y(n)$.

Penggunaan metode ini tergantung dari sinyal yang akan digunakan. Metode Autokorelasi lebih cocok untuk sinyal-sinyal stasioner sedang metode Kovarian lebih sesuai untuk sinyal-sinyal non stasioner.

2.9. Metode Ekstraksi Frekuensi Dasar dan Periode Pitch

2.9.1. Metode Autokorelasi

Estimasi periode pitch dilakukan dengan metode autokorelasi. Deretan Autokorelasi (r_n) dihitung dari keluaran filter invers seperti telah disebutkan pada bab 2.6, keluaran invers filter $\{y(n)\}$, dinyatakan dengan persamaan :

$$y(n) = w(n) + \sum_{i=1}^4 a_i w(n-i), \quad n = 0, 1, 2, \dots, n-1 \quad [2.19]$$

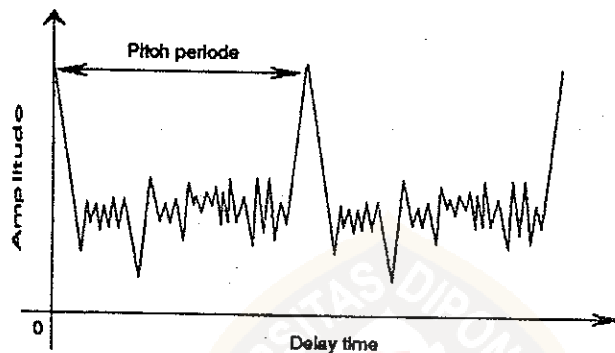
dengan $w(n) = 0, \quad n < 0$

Periode Pitch dapat diestimasi dengan menghitung Autokorelasi untuk m titik data, dimana $m = 2N$ dan N adalah jumlah data dalam satu frame. Persamaan autokorelasi dinyatakan dengan persamaan :

$$r_n = \begin{cases} \sum_{j=0}^{N-1-n} y(j) y(j+n), & n = 0, 1, 2, \dots, m/2-1 \\ 0 & n = m/2 \\ r_{m-n} & n = m/2+1, m/2+2, m-1 \end{cases} \quad [2.20]$$

Setelah deretan nilai Autokorelasi r_n dan keluaran invers filter diperoleh, kemudian nilai-nilai tersebut dinormalisasikan terhadap nilai r_0 . Selanjutnya dicari nilai puncak tertinggi selain r_0 . Letak nilai puncak akan

menentukan besar periode pitch, sedang kebalikan dari periode pitch menunjukkan besarnya frekuensi dasar. Hal tersebut dapat dilihat pada gambar 2.11. yang menunjukkan bagaimana cara menentukan periode pitch berdasar prinsip autokorelasi.



gambar 2.11.

Ekstraksi periode Pitch menggunakan prinsip Autokorelasi

2.9.2. INTERPOLASI

Pada pengukuran frekuensi dasar, diharapkan hasil yang seteliti mungkin. Sedang pada Algoritma SIFT selalu diestimasi lebih dahulu sebelum data diolah, sehingga data yang diolah lebih sedikit. Akibatnya nilai puncak yang diperoleh bukan merupakan nilai puncak yang sesungguhnya (letak informasi periode pitch yang sesungguhnya). Dengan kata lain hasil tersebut kurang akurat. Oleh sebab itu perlu dilakukan interpolasi disekitar titik data nilai puncak agar puncak yang sebenarnya ditemukan. Untuk mencari titik data yang merupakan nilai puncak sebenarnya

tidak mudah karena fungsi sinyalnya tidak diketahui, sehingga nilai fungsinya tidak dapat dihitung, akan tetapi hal ini dapat diatasi dengan pendekatan fungsi (yaitu fungsi sinyal didekati dengan suatu fungsi tertentu).

Dalam periode numerik, jika diketahui m titik data $\{ (x_i, f_i) \}$ untuk $i = 1, 2, \dots, m$ dimana persamaan fungsi $f(x)$ tidak diketahui untuk fungsi $f(x)$ dapat didekati dengan suatu fungsi $g(x)$, sehingga $g(x)$ memenuhi syarat :

$$g(x_i) = f_i ; \quad 1 \leq i \leq m$$

syarat ini merupakan syarat interpolasi, karena fungsi $f(x)$ didekati dengan fungsi $g(x)$, maka dapat dicari nilai fungsi disatu titik manapun dalam interval $1 \leq i \leq m$.

Fungsi $g(x)$ dapat berupa polinomial atau deret Fourier. Bila fungsi $f(x)$ didekati dengan fungsi periodik maka digunakan deret Fourier. Sebagai contoh fungsi periodik yang paling terkenal adalah fungsi Trigonometri. Jika fungsi Trigonometri yang digunakan maka Interpolasinya disebut dengan Interpolasi Trigonometri.

Karena pada Tugas Akhir ini penentuan *Voiced* dan *Unvoiced* didasarkan atas struktur keperiodikan sinyal dan pada dasarnya suatu fungsi dapat dinyatakan dengan penjumlahan fungsi-fungsi periodik, maka interpolasi trigonometri dapat digunakan untuk estimasi periode Pitch dengan seteliti mungkin. Karena fungsi yang digunakan untuk pendekatan adalah fungsi periodik, maka dilakukan pendekatan dengan Fourier.

2.10. ALGORITMA PENENTU VOICED/UNVOICED SINYAL WICARA

Pada Tugas Akhir ini penentuan *Voiced / Unvoiced* dilakukan untuk tiap frame data. Mula-mula nilai puncak hasil interpolasi pada frame k dibandingkan dengan nilai ambang tertentu. Nilai ambang ini dibuat variabel terhadap lokasi puncak dengan patokan satu lokasi yang tetap. Jika lokasi puncak makin kecil, maka nilai ambang dinaikkan karena periode Pitch dalam satu framenya semakin banyak sehingga nilai autokorelasinya makin tinggi. Sebaliknya jika lokasi puncak makin besar, nilai ambang mengecil karena jumlah periode pitchnya semakin sedikit. Jika nilai puncak lebih besar dari nilai ambang, maka lokasi puncak tersebut adalah periode Pitch dan framenya *Voiced*. Sebaliknya bila nilai puncaknya tidak melampau nilai ambang, maka framenya *Unvoiced* dan periode Pitchnya sama dengan nol.

Estimasi periode pitch dilakukan pada jangkau 2,5 ms sampai 15,5 ms. Nilai ambang variabel ditentukan dengan lokasi tetap 9 ms dan mempunyai nilai ambang 0,35. Pada lokasi puncak 2,5 ms, nilai puncak harus melampaui 0,7 sedangkan pada lokasi puncak 15,5 ms nilai ambangnya sebesar 0,315. Penerapan nilai ambang variabel ini dilakukan dengan membagi nilai puncak pada frame ke-k dengan faktor D yang mempunyai variabel lokasi puncak. Untuk Lokasi puncak ≥ 18 (pada sampel ke-18 atau 9 ms), nilai faktor D adalah :

$$D = - 0,1 * (L - 18) / 13 + 1$$

sedangkan untuk lokasi puncak < 18 , nilai faktor D adalah :

$$D = -(L-5)/13 + 2$$

dan $V' = V/D$

dengan : L = lokasi puncak

V = nilai puncak hasil interpolasi

V' = nilai puncak setelah dibagi dengan faktor D

Pada langkah selanjutnya nilai V' dibandingkan dengan nilai ambang tetap 0,35. Jika lebih besar berarti framenya *Voiced*, tetapi bila lebih kecil maka harus dilihat dahulu frame sebelumnya. Jika frame sebelumnya *Voiced*, nilai ambang diturunkan menjadi 0,3, maka framenya *Voiced*, tetapi jika nilainya tetap lebih kecil berarti framenya adalah *Unvoiced*.

Seperti telah disebutkan pada bab sebelumnya, sinyal wicara bersifat stasioner untuk selang waktu yang pendek, yaitu antara 10 - 30 ms. artinya bentuk sinyal wicara hampir tidak berubah untuk selang waktu tersebut. Oleh sebab itu jika frame k dan frame (k-2) *Voiced*, maka dapat dipastikan bahwa frame (k-1) *Voiced* juga, tetapi kadang-kadang timbul kesalahan dimana frame (k-1) dinyatakan *Unvoiced*. Oleh sebab itu pada algoritma penentu *Voiced /Unvoiced* ini juga dilakukan proses koreksi, yaitu dengan melihat nilai periode Pitch dari frame k dan frame (k-2). Jika besarnya tidak jauh berbeda, maka periode Pitch untuk frame (k-1) adalah nilai rata-rata dari periode Pitch frame k dan frame (k-2) serta frame (k-1) dikoreksi menjadi *Voiced*.