

PENGENALAN *VOICED* DAN *UNVOICED* DENGAN ANALISIS *PITCH*

Anni Yuliasuti

L2F 300 501

Jurusan Teknik Elektro Fakultas Universitas Diponegoro
Jl. Prof. Sudarto, Tembalang, Semarang (024) 7460057

ABSTRAK

Sebuah kata yang kita ucapkan merupakan satuan bahasa yang memiliki satu pengertian yang terdiri dari vokal dan konsonan. Apabila kita menganalisis kata berdasarkan penulisan, dengan mudah kita dapat mengenali vokal dan konsonan, tetapi jika kita menganalisis berdasarkan penyuaran sangat sulit untuk mengenalinya, karena konsonan dalam pengucapannya selalu diikuti dengan vokal. Contohnya konsonan “k” diucapkan “ka”. Sehingga sebagai langkah awal untuk pengenalan, lebih mudah dengan membagi dalam dua golongan, yaitu *voiced* dan *unvoiced*.

Metoda yang digunakan untuk pengenalan *voiced* dan *unvoiced* dengan menggunakan analisis *pitch*. *Pitch* adalah hasil akustik dari kecepatan getaran pita suara, sehingga *pitch* hanya berlaku untuk *voiced*, dimana metode untuk mencari nilai *pitch* dengan menggunakan analisis *cepstrum*.

Dengan sampel suara pengucapan beberapa kata dalam bahasa Indonesia dan dibingkai setiap 40 milidetik dihasilkan bahwa *voiced* mempunyai nilai *pitch* lebih besar dari 94.1176 Hz dan *unvoiced* mempunyai nilai *pitch* sama dengan nol Hz. Selain itu semua vokal dikenali sebagai *voiced*, sedangkan konsonan dikenali sebagai *unvoiced*, tetapi ada konsonan yang dikenali *voiced*, hal ini disebabkan karena pengucapan konsonan selalu diikuti vokal.

Kata kunci : *Pitch*, *voiced*, *unvoiced*

I. PENDAHULUAN

Sebuah kata yang kita ucapkan merupakan satuan bahasa yang memiliki satu pengertian yang terdiri dari vokal dan konsonan. Apabila kita menganalisis berdasarkan penulisan, dengan mudah kita dapat mengenalinya, tetapi jika kita menganalisis berdasarkan penyuaran sangat sulit untuk mengenalinya, karena konsonan dalam pengucapannya selalu diikuti dengan vokal. Contohnya konsonan “k” diucapkan “ka”. Sehingga sebagai langkah awal untuk pengenalan, lebih mudah dengan membagi dalam dua golongan, yaitu *voiced* (berupa vokal) dan *unvoiced* (berupa konsonan).

Voiced merupakan hasil getaran dari hambatan pada pita suara ketika arus udara melewatinya. Sedangkan *unvoiced*, terjadi jika tidak ada getaran pada pita suara. Salah satu ciri *voiced* adalah *pitch* yaitu hasil akustik dari kecepatan getaran pita suara. Sehingga untuk pengenalan *voiced* dan *unvoiced* dengan menggunakan metode

analisis *pitch*, dimana untuk mencari nilai *pitch* dengan analisis *cepstrum*

Tujuan penyusunan tugas akhir ini adalah untuk pengenalan sinyal ucapan *voiced* dan *unvoiced* dengan menggunakan analisis *pitch*.

Batasan masalah pada tugas akhir ini, adalah :

1. Menjelaskan dan membedakan *voiced* dan *unvoiced*.
2. Sampel suara menggunakan kata-kata dalam bahasa Indonesia, dimana *voiced* (penyuaran berupa vokal) dan *unvoiced* (penyuaran berupa konsonan).
3. Metoda yang digunakan untuk pengenalan *voiced* dan *unvoiced* dengan menggunakan analisis *pitch*.
4. Membuat simulasi dengan menggunakan program Matlab 5.3.

II. DASAR TEORI

2.1 Proses Terbentuknya Suara

Gambar 1 Alat ucap manusia ^[2]

Keterangan :

1. Paru-paru
2. Batang tenggorok
3. Pangkal tenggorok
4. Pita suara
5. Krikoid
6. Tiroid
7. Aritenoid
8. Dinding rongga kerongkongan
9. Epiglotis
10. Akar lidah
11. Pangkal lidah
12. Tengah lidah
13. Daun lidah
14. Ujung lidah
15. Anak tekak
16. Langit-langit lunak
17. Langit-langit keras
18. Gusi
19. Gigi atas
20. Gigi bawah
21. Bibir atas
22. Bibir bawah
23. Mulut
24. Rongga mulut
25. Rongga hidung

Terjadinya bunyi ^[2] pada umumnya dimulai dengan proses pemompaan udara keluar dari paru-paru menuju keatas dan keluar melalui batang tenggorokan mengalir melalui glotis. Berhadapan dengan glotis terdapat dua selaput yang disebut pita suara, meskipun bentuknya menyerupai tirai. Pita suara ini dapat dirapatkan sama sekali, sehingga menyumbat aliran udara yang keluar (atau masuk), atau diregangkan, ketika udara bisa lewat dengan bebas tanpa mengeluarkan bunyi. Pita suara bisa juga dirapatkan sehingga udara menggetarkannya secara teratur pada kecepatan yang berbeda-beda sewaktu udara itu memaksa melewati pita suara tersebut. Getaran ini secara teknis disebut bersuara (voiced), yang juga merupakan sumber penting dari tinggi nada (pitch). Getaran ini dapat dirasakan dari luar dengan menyentuh bagian depan laring atau jakun dengan ujung jari ketika mendengarkan atau mengucapkan *Ah*. Jika tidak ada getaran disebut tak bersuara (unvoiced).

2.2 Klasifikasi Bunyi

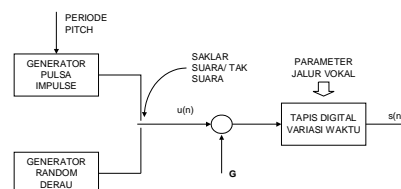
Pada umumnya bunyi ^[2] pertama-tama dibedakan atas vokal dan konsonan. Bunyi vokal dihasilkan dengan pita suara terbuka

sedikit. Pita suara yang terbuka sedikit ini menjadi bergetar ketika dilalui arus udara yang dipompakan dari paru-paru. Selanjutnya arus udara itu keluar melalui rongga mulut yang berbentuk tertentu sesuai dengan jenis vokal yang dihasilkan. Bunyi konsonan terjadi, setelah arus udara melewati pita suara yang terbuka sedikit atau agak lebar, diteruskan ke rongga mulut atau hidung dengan mendapat hambatan di tempat-tempat artikulasi tertentu. Jadi, beda terjadinya bunyi vokal dan konsonan adalah; arus udara dalam pembentukan bunyi vokal, setelah melewati pita suara, tidak mendapat hambatan apa-apa; sedangkan dalam pembentukan bunyi konsonan arus udara itu masih mendapat hambatan atau gangguan. Bunyi konsonan ada yang bersuara ada yang tidak. Yang bersuara terjadi apabila pita suara terbuka sedikit, dan yang tidak bersuara apabila pita suara terbuka agak lebar. Bunyi vokal, semuanya adalah bersuara, sebab dihasilkan dengan pita suara terbuka sedikit.

2.3 Pitch (F0)

Salah satu parameter dari sinyal suara adalah frekuensi fundamental. Frekuensi fundamental dalam istilah instrumen musik dikenal sebagai *pitch* atau nilai frekuensi dari suatu jenis nada. Pitch atau tinggi nada adalah hasil akustik dari kecepatan pita suara. Semakin cepat getaran pita suara, semakin tinggi, tinggi nadanya. Begitu pula sebaliknya. Sehingga *pitch* ini dapat digunakan sebagai ciri bersuara, dimana metode yang digunakan dengan menggunakan analisis cepstrum

Dengan melihat pemodelan untuk sintesa sinyal suara pada gambar 2, generator pulsa impulse memberikan sumber pembangkitan untuk sinyal bersuara berupa fonem vokal (*a/e/i/o/u*) yang dapat diatur selang waktunya oleh parameter-parameter periode *pitch*. Dan generator random derau berfungsi sebagai sumber pembangkitan untuk sinyal tak bersuara.



Gambar 2 Model sintesa suara ^[10]

Jika dianggap model berada pada interval waktu tertentu, sinyal ucapan diasumsikan sebagai hasil dari konvolusi 2 buah sinyal yaitu konvolusi dari respon impuls jalur vokal dan sinyal generator pulsa impulse (untuk sinyal suara ucapan).

Dari parameter klasifikasi sinyal suara dengan periode *pitch* inilah maka analisa untuk mencari nilai *pitch* dapat dilakukan. Salah satu metode untuk mencari nilai *pitch* adalah menggunakan analisa cepstrum.

Konsep dari analisa cepstrum adalah perhitungan melalui metode Real Cepstrum (RC), RC pada sinyal suara $s(n)$, didefinisikan sebagai :

$$c_s(n) = \text{IDFT} \left\{ \log \left| \text{DFT} \{ s(n) \} \right| \right\} \quad (1)$$

dengan menganggap bahwa DFT dari $s(n)$ adalah $S(\omega)$, maka persamaan menjadi :

$$c_s(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |S(\omega)| e^{j\omega n} d\omega \quad (2)$$

dimana n = data sampling untuk deret genap.

$C_s(\omega) = \log |S(\omega)|$ adalah real dan genap.

Untuk proses perhitungan dari RC dalam dilihat pada gambar berikut ini :



Gambar 3 Blok Perhitungan Real Cepstrum [10]

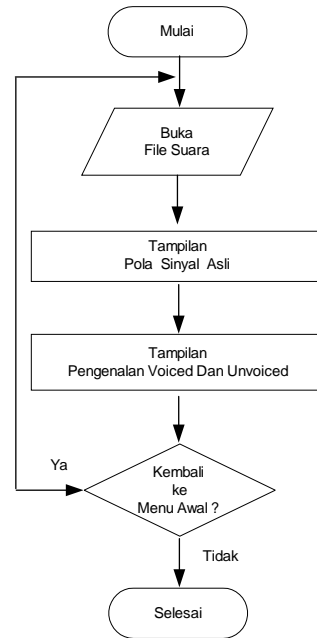
Sinyal suara $s(n)$, melalui penjendeleaan data yang ada, di-Transformasi Fourier-kan menghasilkan cepstrum, kemudian di-log-kan dan di-inverse menghasilkan cepstrum $c_s(n)$. Sehingga cepstrum $c_s(n)$ atau real cepstrum merupakan invers transformasi fourier dari logaritma magnitudo transformasi fourier.

III. PERANCANGAN PROGRAM

3.1 Diagram Alir Perancangan

Pada Tugas Akhir ini, sample suara diambil dengan merekam suara lewat mikrofon yang masuk melalui kartu suara dengan spesifikasi pengaturan frekuensi sampling dan jumlah bit yang disesuaikan kemudian diubah menjadi data digital yang telah disimpan dalam bentuk file wav.

Adapun diagram alir perancangan dari Tugas Akhir ini dapat dilihat pada Gambar 4 sebagai berikut :



Gambar 4 Diagram Alir Pengenalan Voiced dan Unvoiced dengan Analisis Pitch

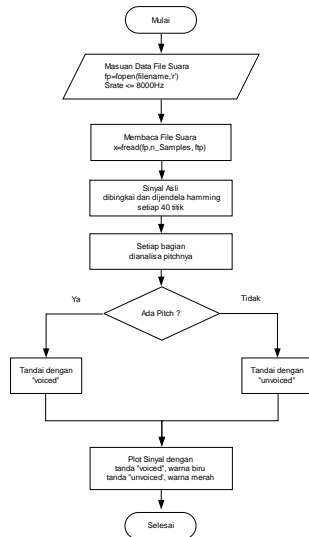
Dalam simulasi pengenalan *voiced* dan *unvoiced* dengan analisis *pitch*, tampilan utamanya dirancang menjadi dua buah menu tampilan yaitu:

1. Menu Tampilan Sinyal Asli.
Akan menampilkan pola sinyal asli suara yang berada dalam kawasan waktu.
2. Menu Tampilan Pengenalan *Voiced* (biru) dan *Unvoiced* (merah).
Akan menampilkan pola sinyal suara sesuai dengan pengenalan *voiced* (biru) dan *unvoiced* (merah) dalam kawasan waktu.

Selain dua tampilan utama tersebut, didukung pula oleh beberapa tombol fungsi untuk kelanjutan dari analisis ini.

3.2 Menu Tampilan Pengenalan Voiced (biru) dan Unvoiced (merah)

Proses tampilan plot pengenalan voiced dan unvoiced dapat dilihat pada diagram alir gambar 5 :



Gambar 5 Diagram alir menu tampilan Pengenalan *voiced* dan *unvoiced*

Dan penggalan senarai programnya sebagai berikut :

```

updRate=floor(40*sr/1000);
fRate=floor(40*sr/1000);
nFrames=floor(n_samples/updRate)-1;

for t=nFrames:-1:1;
    yin=x(t.*fRate:(t.*fRate+fRate-1));
    a=pitch1(fRate, sr, yin)
    if a > 0
        plot(xax(1:t.*fRate), x(1:t.*fRate), 'b')
    else
        plot(xax(1:t.*fRate), x(1:t.*fRate), 'r')
    end
    hold on;
end
hold off;
  
```

Setelah sinyal asli diperoleh, kemudian dibingkai (frame) dan dijendela hamming (*window hamming*) setiap 40 titik. Hasil setiap bagian dianalisa pitchnya. Jika ada *pitch*, plot sinyal tersebut dengan warna

biru. Jika tidak ada, maka plot sinyal tersebut dengan warna merah.

Hasil perhitungan *pitch* itu sendiri merupakan salah satu aplikasi dari metode analisis *cepstrum* sinyal suara. Senarai program fungsi perhitungan *pitch* dengan metode real cepstrum dapat dilihat berikut ini

```

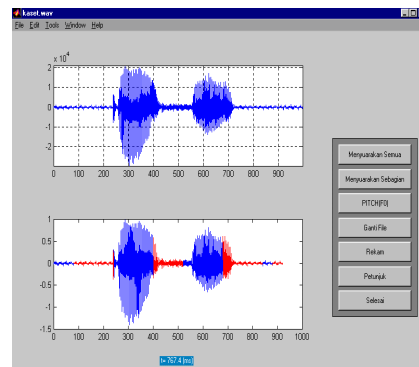
xin=hamming(len).*xin;
cn1=rceps(xin);
LF=floor(sr/500);
HF=floor(sr/70);

cn=cn1(LF:HF);
[mx_cep ind]=max(cn);

if mx_cep > 0.09 & ind > LF
    f0= sr/(LF+ind);
    disp('voiced')
else
    f0=0;
    disp('unvoiced')
end
  
```

IV. HASIL SIMULASI

4.1 Tampilan Simulasi Program

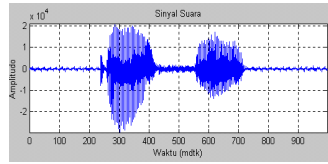


Gambar 6 Tampilan Utama

Gambar 6 menunjukkan tampilan utama program pengenalan *voiced* dan *unvoiced* dengan analisis *pitch*.

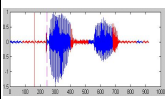
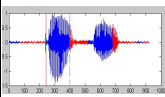
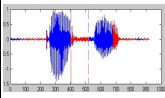
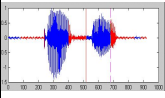
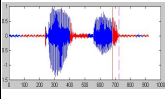
Pada hasil pengujian diamati adanya nilai *pitch* pada setiap *frame*. Jika ada nilai *pitch* maka disebut *voiced* dan ditandai dengan warna biru. Tapi jika nilai *pitch*nya nol (0) Hz maka disebut *unvoiced* dan ditandai dengan warna merah.

4.2 Analisa dan Hasil Pengujian dengan sample suara “kaset”



Gambar 7 Tampilan sinyal asli suara “kaset”

Tabel 1 Hasil pengujian dengan sample suara “kaset”

Sinyal	Interval waktu (milidetik)	Nilai pitch (Hz)	Suara	Keterangan
	201 - 240	0	Mewakili konsonan k	Unvoiced
	241 - 280 281 - 320 321 - 360 361 - 400	228.5714 228.5714 228.5714 222.2222	a	Voiced
	401 - 440 441 - 480 481 - 520	0 0 0	Mewakili konsonan s	Unvoiced
	521 - 560 561 - 600 601 - 640 641 - 680	235.2941 242.4242 235.2941 235.2941	e	Voiced
	681 - 720	0	Mewakili konsonan t	Unvoiced

Tabel 1 menunjukkan bahwa sinyal suara “kaset” pada saat interval waktu antara 200 - 240, 401 - 520 dan 681 – 720 milidetik mempunyai nilai *pitch* nol (0) dan menyuarakan mewakili konsonan “k”, “s” dan “t”, sehingga termasuk *unvoiced*. Sedangkan pada interval waktu antara 241 – 400 dan 521 – 680 milidetik mempunyai nilai *pitch* antara 222.2222 – 242.4242 Hz dan menyuarakan vokal “a” dan “e”, sehingga termasuk *voiced*.

V. KESIMPULAN DAN SARAN

Dari hasil pengujian dan analisis pada tugas akhir ini, maka dapat diambil kesimpulan bahwa :

1. Pembesaran setiap 40 milidetik menghasilkan pengenalan yang lebih baik, untuk membedakan *voiced* dan *unvoiced*.
2. Dengan sampel suara mengucapkan beberapa kata dalam bahasa Indonesia dapat dianalisis bahwa *voiced* mempunyai nilai *pitch* lebih besar dari 94.1176 Hz dan *unvoiced* mempunyai nilai *pitch* sama dengan nol Hz.
3. Semua vokal bahasa Indonesia (a, i, u, e, o) merupakan sinyal bersuara (*voiced*), sedangkan konsonan termasuk sinyal tak bersuara (*unvoiced*), tetapi ada konsonan yang dikenali *voiced*, hal ini disebabkan karena pengucapan konsonan selalu diikuti vokal, sehingga sulit untuk mengenalinya. Contohnya konsonan “l”, diucapkan “el”.

Dalam tugas akhir ini disampaikan beberapa saran-saran antara lain :

1. Saat merekam sampel suara harus diperhatikan :
 - a) Cara pengucapannya
Cara pengucapan yang tidak benar akan mempengaruhi dalam pengenalan *voiced* dan *unvoiced*.
 - b) Ruang
Ruang yang tenang akan mengurangi sinyal derau, sehingga hasilnya akan baik.
 - c) Waktu pengucapan
Disesuaikan dengan durasinya yaitu 1 detik, sehingga sampel tidak terpotong.
2. Analisis *pitch* bisa digunakan untuk menganalisis ucapan bersuara dan berbisik. Tetapi kendalanya pada pengambilan sampel suara. Karena pada saat merekam suara berbisik, banyak sinyal derau yang ikut terekam, sehingga menghasilkan nilai *pitch* lebih besar dari ucapan bersuara.

VI DAFTAR PUSTAKA

- [1] Alan.V. Oppenheim, Ronald W. Schaffer, *Discrete-Time Signal Processing*, Prentice hall, New Jersey, 1989.
- [2] Drs. Abdul Chaer, *Linguistik Umum*, PT. RINEKA CIPTA, Jakarta, 1994.
- [3] r.h. robins, *Lingustik Umum sebuah pengantar*, KANISIUS, Yogyakarta, 1992.
- [4] J.W.M. Verhaar, *Pengantar linguistik*, Gadjah Mada University press, 1995.
- [5] Drs. Suhendra Yusuf, M.A, *Fonetik dan Fonologi*, PT. Gramedia Pustaka Utama, Jakarta, 1998.
- [6] Leonard Janar, Juan Jose Bonet, Eduardo Licida – Soleno, *Pitch Detection and Voiced/Unvoiced based on Wavelet Transforms*.
- [7] Philipos C. Loizou, *A Matlab Software Analysis of Speech*, Colea, 1998.
- [8] Duane Hanselman & Bruce Littlefield, *Matlab bahasa Komputasi Teknis*, ANDI, Yogyakarta, 2000.
- [9] Eko Budi S, *Pengenalan Sinyal Suara Manusia dengan Komputer Pribadi*, Teknik Elektro Universitas Diponegoro, Semarang, 1995.
- [10] R. Yudhi Wismono B, *Identifikasi Jenis Tingkatan Suara Manusia dengan Metode Real Cepstrum*, Teknik Elektro Universitas Diponegoro, Semarang, 2002.
- [11] John N Little, Loren Shure, *Signal Processing ToolBox*, The MathWork, Inc, 1993.



ANNI YULIASTUTI,
Lahir di Karanganyar, 08 Desember 1975. Telah menyelesaikan studi di SD Cangakan III, SMPN 1 Karanganyar, SMAN 1 Karanganyar dan DIII POLITEKNIK UNDIP Teknik Elektro. Saat ini sedang menyelesaikan Tugas Akhir sebagai syarat meraih gelar Srata-1 (S-1) di Teknik Elektro Universitas Diponegoro Semarang dengan konsentrasi jurusan Elektronika.

**Menyetujui,
Pembimbing II**

Achmad Hidayatno, ST, MT
NIP. 132 137 933

**Mengetahui,
Pembimbing I**

Sumardi, ST, MT
NIP. 132 125 670