EKSTRAKSI CIRI BUNYI VOKAL INDONESIA

Samiadji Herdjunanto dan Wim T. Tjambolang

Jurusan Teknik Elektro
Fakultas Teknik, Universitas Gadjah Mada
Yogyakarta

INTISARI

Penelitian ini berupaya mencari ciri bunyi vokal Indonesia yang dimiliki oleh sebarang pembicara dalam rangka menuju pengembangan mesin cerdas yang dapat mengenali perintah sebarang pengguna.

Ternyata sangat sulit untuk menentukan ciri bunyi vokal pada kawasan waktu sehingga perlu diadakan transformasi ke kawasan frekuensi. Pada kawasan frekuensi didapatkan suatu parameter suara yang dapat dijadikan ciri, yaitu frekuensi forman. Dari hasil percobaan tampak bahwa forman ciri yang dihasilkan mengelompok pada daerah frekuensi tertentu.

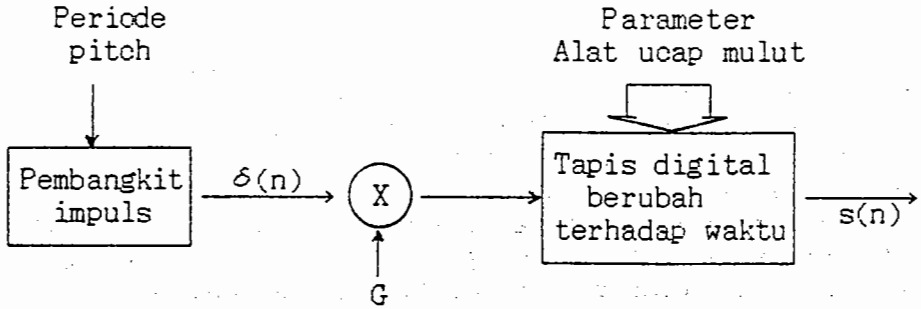
PENGANTAR

Dalam bidang telekomunikasi dewasa ini, pengiriman suara jarak jauh secara digital dilakukan dengan cara pencuplikan dan digitalisasi langsung pada isyarat gelombang suara itu. Laju bit dapat ditekan menjadi cukup rendah dengan menggunakan teknik kompresi dan penyandian data atau yang dikirim adalah parameter pembentuk suara. Penurunan laju bit yang lebih drastis akan diperoleh bila yang dikirim hanya ciri suara dan selanjutnya, berdasarkan ciri itu, penerima akan menyintesiskan suara yang serupa.

Salah satu karakteristik mesin cerdas adalah dapat menangkap perintah pembicara yang beragam. Manfaat mesin ini tampak jelas kalau mesin itu dihubungkan dengan perangkat lain, misalnya mesin ketik atau komputer yang tersambung pada saluran telepon sebagai pusat data jarak jauh dengan tanggapan berupa suara.

Kedua hal tersebut di atas mendorong diadakannya studi tentang ciri suara, yang membantu membedakan suatu suara dengan suara yang lain. Ciri itu juga dapat menentukan suatu jenis suara. Misalnya ciri suara dapat mengelompokkan bunyi vokal /a/ yang diucapkan oleh dua orang yang berbeda, tetapi juga dapat membedakan dua bunyi vokal yang berbeda /a/ dan /i/.

Studi ini bertujuan mencari parameter-parameter suara yang dapat digunakan sebagai ciri untuk membedakan bunyi vokal yang satu dengan yang lain serta mengidentikkan bunyi-bunyi vokal yang sama dari pembicara yang berlainan. Gambar 1 memperlihatkan model pembentukan bunyi vokal yang disederhanakan. Bunyi vokal terbentuk oleh tapis yang dimasuki oleh sederetan gelombang impuls. Penapisan inilah yang menentukan bunyi vokal itu adalah /a/ atau /i/, atau yang lain. Oleh karena itu, setiap bunyi vokal dapat dimodelkan sebagai tapis digital yang berubah dengan waktu. Dengan kata lain, bunyi vokal yang berbeda akan menghasilkan suatu set parameter tapis yang berbeda pula.



Gambar 1. Model pembentukan bunyi vokal

Dari gambar 1 s(n) dapat dinyatakan dengan persamaan (1):

$$s(n) = \sum_{k=1}^{p} \alpha_k \, s(n-k) + G \, \delta(t) \qquad \ldots (1),$$

dengan:

s(n) : cuplikan data suara ke n
s(n-k): cuplikan data suara ke (n-k)
α_k : koefisien tapis pada model
G : parameter penguatan
$\delta(t)$: deretan impuls.

Tapis digital pada model pembentukan bunyi vokal berjenis linear *all pole*. Selanjutnya, dilakukan estimasi linear cuplikan suara ke n dengan menggunakan runtun data suara seperti tampak pada persamaan (2). Estimator ini disebut juga sebagai peramal linear karena estimasi cuplikan suara ke n adalah sebagai linear kombinasi cuplikan-cuplikan suara sebelumnya.

$$\hat{s}(n) = \sum_{k=1}^{p} a_k \, s(n-k) \qquad \ldots (2).$$

Galat peramalan yang terjadi didifinisikan:

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} a_k\, s(n-k) \qquad \ldots (3).$$

Dengan membuat koefisien-koefisien tapis sama dengan koefisien-koefisien estimator, maka diperoleh fungsi alih tapis seperti tampak pada persamaan (4):

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k\, z^{-k}} \qquad \ldots (4).$$

Karena sifat isyarat ucapan berubah dengan waktu, maka koefisien peramal $\{a_k\}$ harus diestimasikan dari segmen-segmen pendek isyarat suara. *Performance index* yang digunakan untuk estimasi linear ini adalah minimisasi rerata kuadrat galat peramalan (*least mean square error).*

Rerata kuadrat galat peramalan untuk suatu segmen pendek didifinisikan dengan menggunakan asumsi bahwa galat peramalan, e(n), orde p akan bernilai tidak nol dalam interval $0 \le n \le N-1+p$, sehingga E dapat dituliskan sebagai:

$$E = \sum_{n=0}^{N+p-1} e^2(n) \qquad \ldots (5).$$

Selanjutnya, dibuat $\delta E/\delta a_i = 0$, untuk i = 1, 2, ..., p, sehingga akan diperoleh koefisien peramal a_k. Hasil penjabaran ini menghasilkan persamaan :

$$\sum_{n=0}^{N+p-1} s(n-i)s(n) = \sum_{k=1}^{p} a_k \sum_{n=0}^{N+p-1} s(n-i)\, s(n-k) \text{ untuk } 1 \le i \le p \qquad \ldots (6).$$

Bila didifinisikan :

$$\phi(i,k) = \sum_{n=0}^{N+p-1} s(n-i)s(n-k) \qquad \ldots (7),$$

maka persamaan (6) dapat ditulis dalam bentuk yang lebih kompak sebagai berikut:

$$\sum_{k=1}^{p} a_k\, \phi(i,k) = \phi(i,0) \quad \text{untuk} \quad i = 1,2,\ldots,p \qquad ..(8).$$

Karena s(n) bernilai nol diluar interval $0 \le n \le N-1$, maka persamaan (7) dapat ditulis kembali sebagai berikut:

$$\phi(i,k) = \sum_{n=0}^{N-1-(i-k)} s(n)\, s(n+i-k) \quad \text{untuk } 1 \le i \le p \text{ dan } 0 \le k \le p \qquad \ldots (9).$$

Selanjutnya, diketahui bahwa *autocorrelation function* suatu isyarat (misal s(n)) adalah :

$$R(m) = \sum_{n=0}^{N-1-m} s(n)\, s(n+m) \qquad \ldots (10).$$

Dengan membandingkan persamaan (9) dengan persamaan (10) dan mengingat bahwa R(m) adalah fungsi genap, maka didapatkan persamaan (11):

$$\sum_{k=1}^{p} a_k\, R(|i-k|) = R(i) \text{ untuk } 1 \le i \le p \qquad \ldots (11).$$

Koefisien $\{a_k\}$ didapat dengan menyelesaikan persamaan (11) dengan algoritma rekursif Durbin (Giordano dan Hsu,1985). Parameter penguatan G untuk tapis digital dapat dihitung dengan persamaan (3) dan persamaan (5). Frekuensi forman adalah nilai-nilai puncak yang terlihat pada tanggapan frekuensi tapis.

CARA PENELITIAN

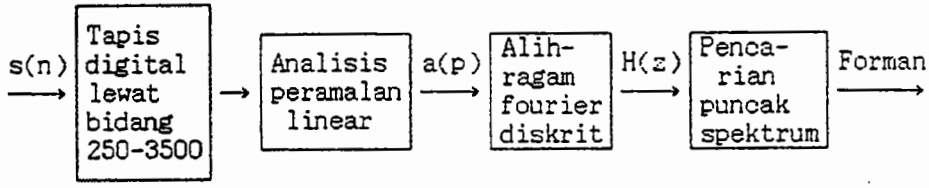
Mula-mula dikumpulkan beberapa sampel suara dengan merekam kata-kata yang berasosiasi dengan bunyi vokal yang diinginkan, dari beberapa pembicara. Kata-kata itu dapat dilihat pada daftar I. Suara selanjutnya didigitalkan dengan menggunakan kartu suara *Sound Blaster* dengan pesat pencuplikan 10 kHz.

Daftar I. Kata yang digunakan sebagai sampel

Vokal :	/a/	/i/	/u/	/e/	/ɛ/	/o/
Kata :	bata	bisa	buta	beda	beban	bobot

Proses selanjutnya adalah pemilihan isyarat vokal dari rekaman suara itu. Pemilihan suara didasari oleh besarnya energi suara. Energi bagian vokal tampak lebih besar dari pada energi bagian konsonan. Karena vokal yang dimaksud berada pada bagian awal kata, maka isyarat vokal diperoleh dengan mencari segmen berenergi tinggi pada awal rekaman.

Satu segmen isyarat vokal mengandung minimal dua gelombang periodis. Lebar segmen adalah 25,6 mdetik, sehingga bisa didapat 256 data. Proses dilanjutkan dengan menapis data melalui tapis digital lewat-bidang antara 250 Hz sampai 3500 Hz. Tapis diskret yang digunakan adalah jenis Butterworth orde 10 (Stearns dan Hush,1990). Selanjutnya, dihitung koefisien-koefisien peramal linear dengan menggunakan metode Durbin. Koefisien-koefisien peramal ini dipakai untuk mendapat tanggapan frekuensi tapis melalui alihragam Fourier. Frekuensi forman diperoleh dengan cara mencari puncak-puncak spektrum tapis yang dihasilkan (gambar 2).



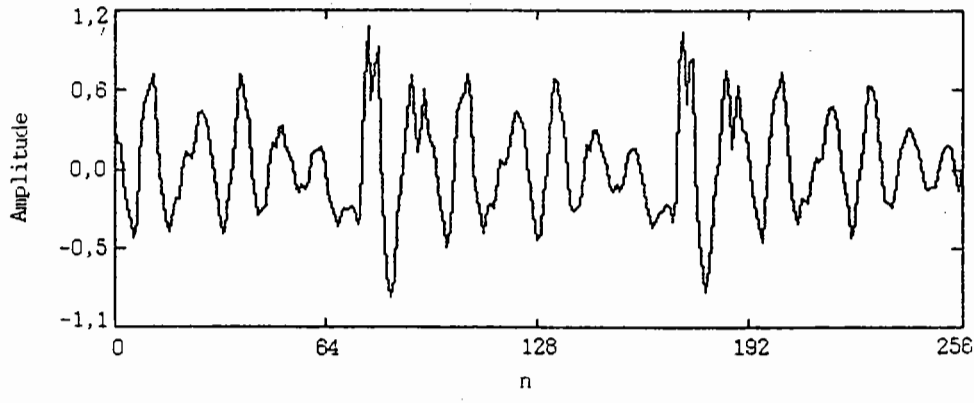
Gambar 2. Blok diagram proses ekstraksi forman

HASIL DAN PEMBAHASAN

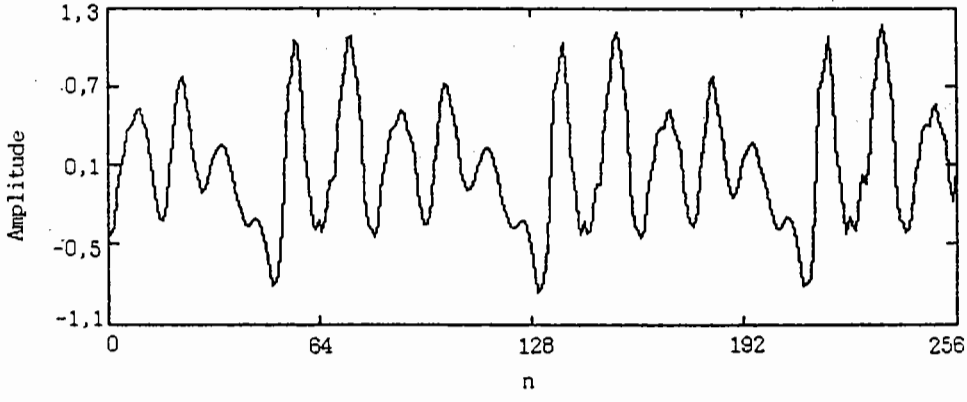
Gambar 3 memperlihatkan bentuk gelombang bunyi vokal /a/ yang diambil dari dua pembicara yang berbeda. Bila diadakan pengamatan terhadap kedua gelombang vokal itu, maka akan sangat sulit mengetahui bahwa kedua bunyi vokal itu merupakan bunyi vokal yang sama. Gambar 4 memperlihatkan keadaan yang serupa namun untuk bunyi vokal /e/.

Setelah diamati pada kawasan frekuensi maka lebih mudah menyimpulkan bahwa kedua vokal itu mempunyai bunyi vokal yang sama. Gambar 5 memperlihatkan spektrum frekuensi tapis yang membentuk bunyi vokal pada gambar 3. Bila diperhatikan nilai-nilai frekuensi forman pada gambar 5a dan gambar 5b yang berkorespondensi dengan isyarat bunyi pada gambar 3a dan gambar 3b, maka keduanya mempunyai nilai yang sangat dekat. Atas dasar inilah dapat disimpulkan bahwa kedua isyarat itu merupakan bunyi vokal yang sama. Gambar 6 memperlihatkan hal yang serupa untuk bunyi vokal /e/.

Frekuensi rata-rata untuk forman ke-1, forman ke-2, dan forman ke-3 untuk enam bunyi vokal yang berlainan diberikan pada daftar II. Gambar 7 merupakan grafik sebaran data antara forman ke-1 dan forman ke-2. Gambar 8 dan gambar 9 berturut-turut memperlihatkan grafik sebaran data antara F1 dan F3 serta antara F2 dan F3.



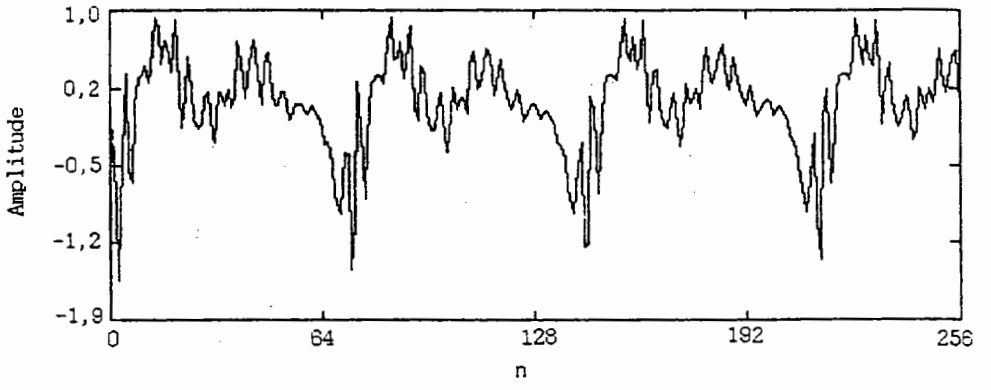
(a)



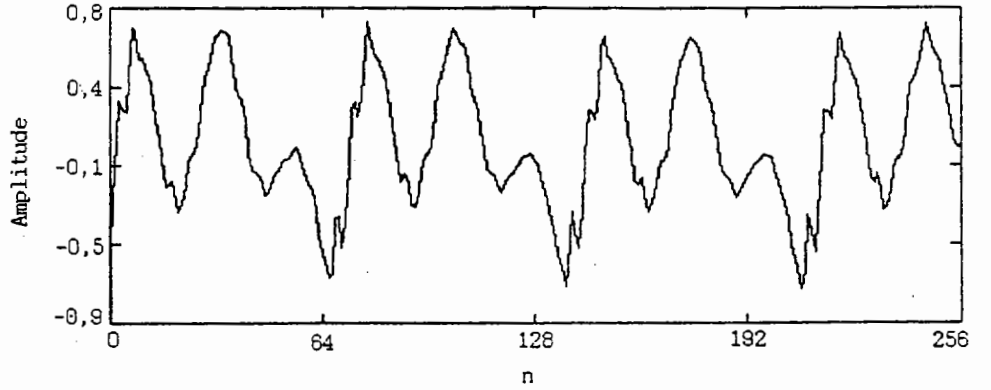
(b)

Gambar 3. Bentuk gelombang bunyi vokal /a/ seperti pada kata /bata/

Dengan mengamati gambar 7 maka dapat dilihat bahwa setiap bunyi vokal terkonsentrasi pada satu kawasan frekuensi tertentu. Walaupun terjadi *overlap* sedikit antara vokal /u/ dan /o/ serta antara vokal /i/ dan /e/, tetapi keduanya tetap cenderung mengelompok dan membentuk kelas tersendiri. Hal yang sama tidak teramati pada gambar 8 dan gambar 9. Pada kedua gambar terakhir pasangan forman tidak cukup baik untuk memisahkan ke enam bunyi vokal itu.



(a)

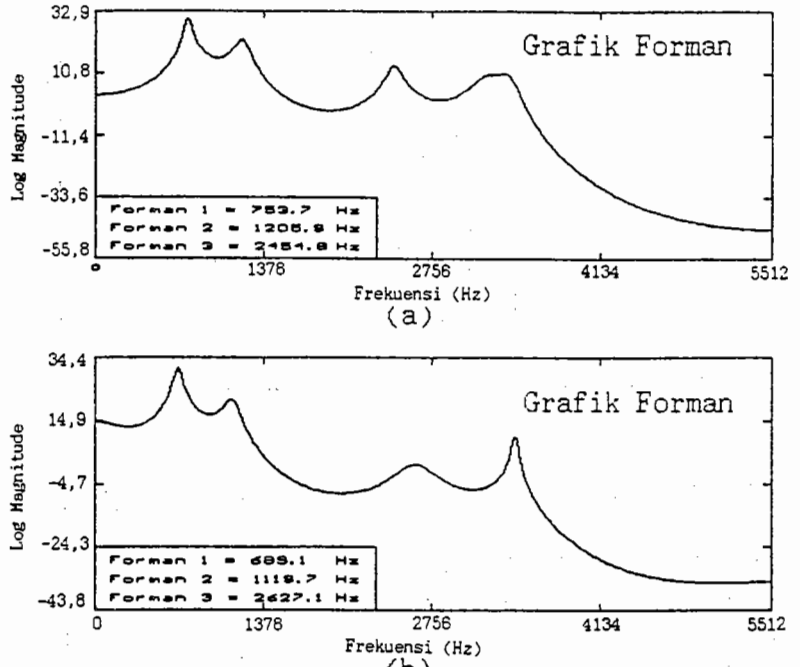


(b)

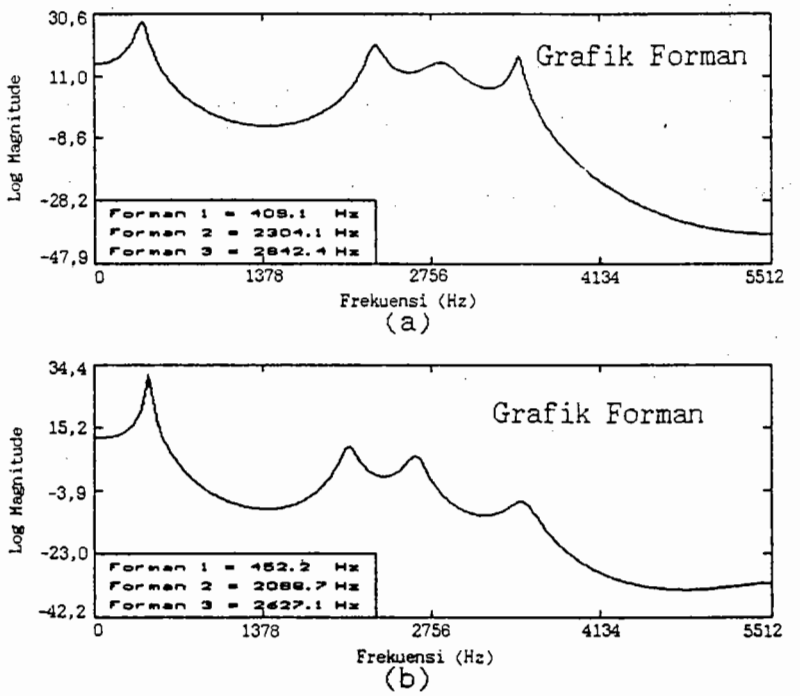
Gambar 4. Bentuk gelombang bunyi vokal /e/ seperti pada kata /beda/

Daftar II. Frekuensi rata-rata forman bunyi vokal

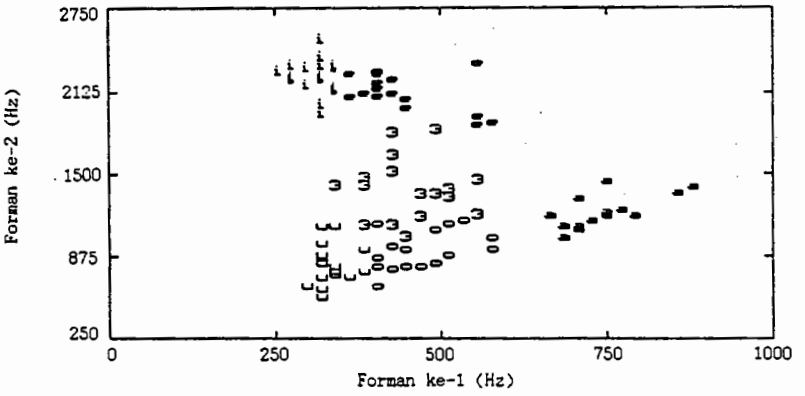
Vokal	F1	F2	F3
/a/	717,5	1236,9	2595,0
/i/	330,2	2269,0	3048,0
/u/	352,1	841,9	2375,1
/e/	427,7	2145,5	2704,0
/ɛ/	441,4	1440,6	2552,8
/o/	458,4	923,9	2527,6



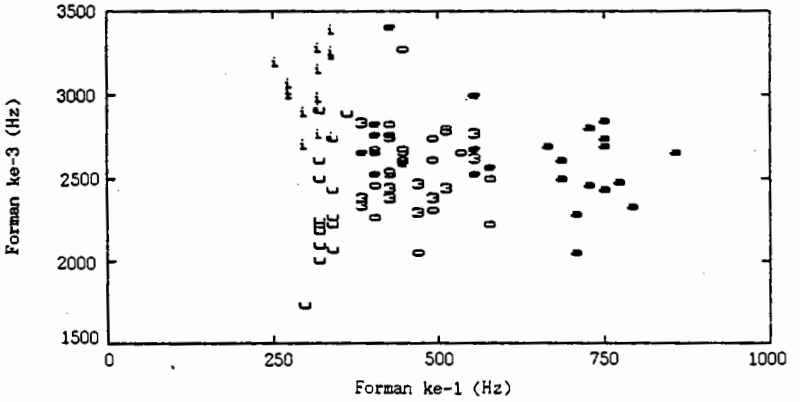
Gambar 5. Spektrum frekuensi tapis untuk bunyi vokal /a/ seperti pada kata /bata/



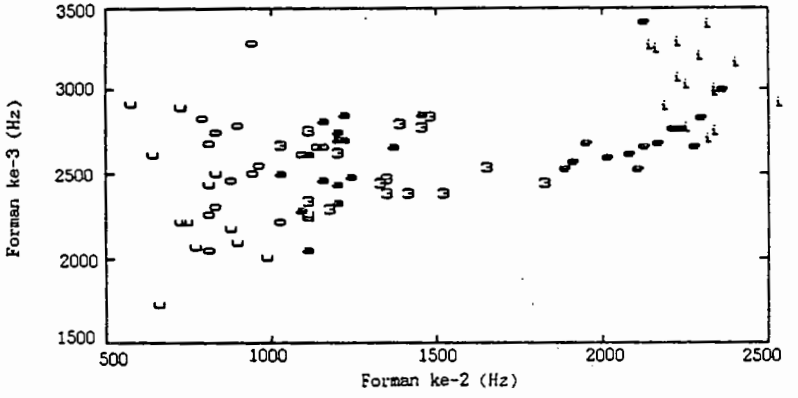
Gambar 6. Spektrum frekuensi tapis untuk bunyi vokal /e/ seperti pada kata /beda/



Gambar 7. Grafik sebaran data antara F1 dan F2



Gambar 8. Grafik sebaran data antara F1 dan F3



Gambar 9. Grafik sebaran data antara F2 dan F3

KESIMPULAN

1. Sangat sulit mencari kesamaan ciri antara bunyi vokal yang sama dalam kawasan waktu.

2. Parameter ciri bunyi vokal bisa didapatkan pada kawasan frekuensi, yaitu frekuensi forman.

3. Ciri bunyi vokal sangat baik dipresentasikan dengan pasangan forman ke-1 dan forman ke-2.

DAFTAR PUSTAKA

Giordano, A.A. and Hsu, F.M., 1985, "Least Square Estimation with Applications to Digital Signal Processing", John Wiley and Sons, New York.

Stearns,S.D. and Hush,D.R., 1990, "Digital Signal Analysis", Prentice Hall Inc, Englewood Cliffs, New Jersey.