

SIMULATOR MODEL LINEAR DINAMIK (KALMAN FILTER) UNTUK PENYELEKSIAN DIPHONE PADA PENSINTESA SUARA BAHASA INDONESIA

Muhammad Subali

Program Diploma Ilmu Komputer, Universitas Gunadarma

Jln Margonda Raya No 100 Pondok Cina Depok

subali@staff.gunadarma.ac.id

ABSTRAK

Sistem pensintesa suara adalah suatu teknik dalam pemrosesan signal suara dimana pada sistem ini suatu teks dapat diubah menjadi signal suara. Teknik diphone concatenative adalah suatu teknik dimana segmen suara dalam bentuk diphone sudah direkam sebelumnya berdasarkan pada kaidah yang berlaku dalam suatu bahasa. Suatu teks akan disintesa menjadi suara berdasarkan pada gabungan diphone penyusun pada teks tersebut. Untuk memperoleh hasil suara yang alami, maka perlu diseleksi (dipilih) diphone mana yang tepat untuk digabungkan. Banyak metode dapat digunakan untuk melakukan pemilihan diphone, dan pada tulisan ini akan digunakan metode model linear dinamik yaitu suatu metode parametrik yang akan memperkirakan parameter dari sistem (diphone). Dimana model ini merupakan sistem pengaturan ruang-keadaan.

Kata kunci : Model linear dinamik, diphone concatenative, sintesa suara, ruang-keadaan

PENDAHULUAN

Penelitian di bidang pensintesa ucapan mengalami perjalanan yang sangat panjang dan telah dimulai sejak lama. Perkembangan teknologi pensintesa ucapan dapat dibagi menjadi tiga kurun waktu. Kurun waktu pertama adalah sebelum 1930. Pada masa ini penelitian lebih banyak ditujukan untuk memahami karakteristik sinyal ucapan serta pengembangan pensintesa ucapan berbasis mekanik atau elektromekanik. Kurun waktu kedua dimulai sejak tahun 1930-an sampai dengan ditemukannya komputer digital. Masa ini ditandai dengan pengembangan berbagai alat pensintesa ucapan menggunakan teknologi elektronik analog. Kurun waktu ketiga dimulai sejak ditemukannya komputer digital hingga sekarang. Pada masa ini, sintesa ucapan dilakukan menggunakan pendekatan pemrosesan sinyal digital.

Berkembangnya komputer digital tidak hanya menyebabkan berkembangnya sistem pensintesa ucapan, tetapi juga melahirkan alternatif baru untuk mengimplementasikan pengisian pembangkit ucapannya. Pada era komputer digital, pembangkitan ucapan dilakukan menggunakan algoritma pemrosesan sinyal digital yang diimplementasikan menggunakan perangkat lunak. Bentuk pensintesa digital yang berkembang pada awalnya adalah pensintesa yang dikenal

dengan istilah *formant synthesizer*, yang bekerja dengan cara mensimulasikan kompone frekuensi utama pembentuk ucapan yang disebut formant

Salah satu masalah dalam teknik *concatenative* adalah terjadinya suara yang terdengar secara diskontinu seputar diphone tersebut terutama yang menonjol untuk fonem dalam kelompok vokal dan semi-vokal. Sehingga untuk menangani masalah ini dapat pula diatasi menggunakan unit yang lebih besar misalnya triphone, namun hal ini akan menyebabkan penambahan secara drastis pada inventori (basis data).

Untuk Pesintesa teknik non parametrik (OLA,PSOLA, MBROLA,TD-PSOLA dan lain-lain) suara yang dihasilkan masih terdapat adanya diskontinuitas atau terdapat bunyi fonem yang tidak terdengar. Hal ini disebabkan karena adanya ketidaksesuaian fase, spektral dan pitch (Dutoit dan Leich, 93,96). Selain itu teknik ini juga masih belum dapat menghasilkan hasil sintesa dengan karakteristik bunyi yang bervariasi seperti gaya bicara, emosi dan sebagainya. Pada teknik *concatenative* hal ini dapat dicapai namun memerlukan data suara dengan jumlah yang besar.

PEMBAHASAN

Pada pemilihan unit bunyi diphone 'ai' dan 'ay' untuk suku kata yang mengandung

diftong 'ai' mempunyai tiga (3) buah input. Pertama, input suku kata dimana signal suara dari kata yang digunakan telah direkam. Untuk input pada signal suara suku kata hanya diambil diphone terakhir pada kata yang dipilih yaitu diphone 'ai,' jadi tidak semua signal suara diphone pada suku kata yang dipilih dimasukkan. Kedua, input diphone 'ai' yaitu berupa signal suara pengucapan kata 'ai' yang telah direkam. Ketiga, input diphone 'ay' juga berupa signal suara pengucapan kata 'ay'.

Proses pemilihan diphone selanjutnya yaitu proses segmentasi. Signal suara yang dimasukkan pada masing-masing input mempunyai kisaran panjang gelombang yang berbeda-beda, sehingga diperlukan suatu batasan kisaran panjang gelombang dari setiap signal suara input, proses segmentasi ini yang akan menentukan berapa kisaran panjang gelombang yang dipakai untuk setiap signal suara input. Setelah didapat nilai kisaran maka setiap diphone input akan dibandingkan dengan suku kata yang dipilih sebagai input. Hasil dari perbandingan tersebut akan didapat nilai derau untuk masing-masing diphone. Signal derau tersebut digunakan untuk menghasilkan nilai rata-rata daya untuk setiap diphone, dengan kata lain signal derau ini dipakai untuk membentuk r_1 untuk diphone 1 dan r_2 untuk diphone 2.

Proses selanjutnya ialah transformasi Fourier. Signal suara yang akan ditransformasi Fourier terdiri dari signal suara diphone1 yaitu 'ai', signal suara diphone2 yaitu 'ay', signal derau diphone1, serta signal derau diphone2. Dari transformasi Fourier akan didapat nilai nyata dan imajiner masing-masing diphone dan signal derau tetapi nilai yang digunakan untuk proses selanjutnya ialah nilai nyata dari setiap input. Transformasi Fourier menghasilkan spektrum untuk signal diphone1, diphone2, dan signal derau untuk setiap diphone. Nilai yang dihasilkan dari transformasi Fourier merupakan spektrum dari setiap signal yang ditransformasi. GUI pada pemilihan diphone ini hanya menampilkan grafik spektrum untuk signal diphone1 yaitu 'ai' dan diphone2 yaitu 'ay'

Tahap selanjutnya ialah proses perkiraan yang menggunakan parameter model AR, dimana model AR ini akan mengambil

komponen nyata dari spektrum pada signal diphone1, diphone2, dan signal derau untuk setiap diphone. Proses perkiraan tidak hanya menggunakan parameter model AR tetapi bisa menggunakan parameter lain. Hasil perkiraan akan membentuk metriks ruang keadaan. Metriks yang terbentuk berisi nilai dari komponen nyata untuk masing-masing diphone. Nilai tersebut disubstitusikan ke dalam matriks O yang berukuran 16×16 , hasil perkiraan ini akan didapat metriks ruang keadaan A , C , dan Q , dimana metriks ini akan digunakan untuk proses penyaring Kalman.

Proses penyaring Kalman yang dijalankan terdiri dari suku kata yang dipilih, metriks A , metriks C , metriks Q , dan nilai rata-rata daya. Output diphone1 dari penyaring Kalman adalah $Y1$ dan diphone2 adalah $Y2$. Signal keluaran ditampilkan pada tampilan GUI dimana grafik tersebut menggambarkan dari perbandingan antara karakteristik antara diphone dengan diphone tetapi dalam bentuk metriks ruang keadaan. Output $Y1$ dan $Y2$ akan dibandingkan dengan masing-masing diphone, dari dicari nilai rata-ratanya. Diphone yang mempunyai nilai rata-rata terkecil akan diambil sebagai pembentuk ucapan untuk suku kata yang dipilih.

Blok Input

Blok input ditunjukkan Gambar 1. Blok input terdiri dari memasukkan suku kata yang mengandung diftong 'ai' dan diphone 'ai' dan 'ay.' Suku kata yang digunakan pada program simulasi pemilihan diphone ini terdiri dari kata 'pandai', 'ramai', 'tunai', 'tangkai', 'lunglai', 'perisai', dan 'gapai'. Suku kata dan diphone yang dimasukkan berupa signal suara yang telah direkam dalam bentuk WAV.

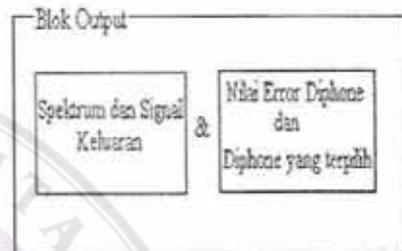
Kata yang mengandung diftong 'ai' disimpan pada basis data dalam bentuk WAV. Jika ingin menggunakan kata-kata tersebut, tinggal memanggil kata-kata tersebut, pada tampilan GUI untuk input suku kata tertentu tinggal dipilih pada menu *popup*.



Gambar 1. Blok Input

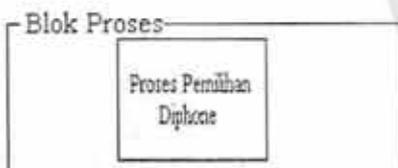
Grafik spektrum diphone1 pada tampilan GUI merupakan gambaran dari nilai signal suara 'ai' sedangkan untuk grafik spektrum diphone2 merupakan gambaran dari nilai signal suara 'ay.' Signal keluaran menggambarkan nilai dari hasil proses penyaring Kalman. Proses penyaring Kalman dapat dilihat pada Gambar 3. Berdasarkan gambar tersebut, input yang digunakan ialah suku kata yang mengandung diftong 'ai,' dan matriks ruang keadaan yang terdiri dari matriks A, C, dan Q, serta rata-rata daya yang terbentuk dari signal derau untuk masing-masing diphone input.

Diphone yang digunakan pada simulasi ini hanya diphone 'ai' dan 'ay' walaupun banyak lagi diphone dalam bahasa Indonesia yang dapat digunakan. Sama halnya dengan suku kata diphone 'ai' dan 'ay' telah tersimpan pada basis data dalam format WAV. Pada tampilan GUI program simulasi ini diphone tersebut bisa langsung diakses di dalam menu popup.



Gambar 3. Blok Output

Proses



Gambar 2 Blok Proses

Proses pemilihan diphone dilakukan dalam beberapa tahap, yaitu segmentasi signal input, transformasi Fourier signal input, estimasi parameter signal input menggunakan model AR, penentuan matriks ruang keadaan, dan proses penyaring Kalman. Tahapan tersebut dilakukan secara berurut karena hasil dari setiap tahap sebelumnya mempengaruhi proses selanjutnya.

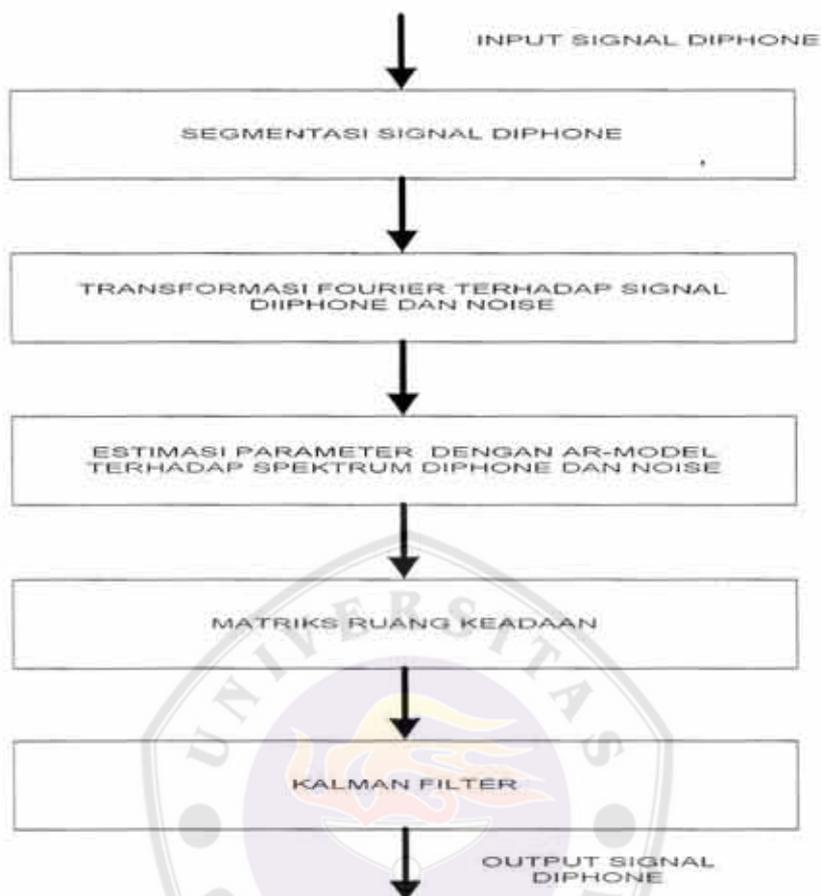
Output

Grafik yang ditampilkan pada GUI untuk spektrum diphone ialah tampilan grafik yang dihasilkan setelah proses perkiraan. Grafik tersebut merupakan gambaran dari nilai nyata yang dihasilkan setelah transformasi Fourier.

Nilai kesalahan yang ditampilkan pada tampilan GUI ini merupakan keseluruhan nilai yang dihasilkan mulai dari proses perkiraan hingga proses penyaring Kalman. Nilai kesalahan ini mengindikasikan diphone mana yang terpilih.

Algoritma pemilihan diphone 'ai' dan 'ay' untuk suku kata yang mengandung diftong 'ai' dapat dilihat pada Gambar 4. Daftar program pada proses input signal diphone, terdiri dari program untuk diphone1, diphone2, input suku kata, seperti yang dapat dilihat pada Tabel 1 dan 2.

Panjang gelombang dari signal suara yang dimasukkan tidak mempunyai panjang yang sama sedangkan jika membandingkan suatu signal tertentu maka kedua signal tersebut harus mempunyai nilai panjang gelombang yang sama. Oleh karena itu, fungsi dari segmentasi signal diphone ialah memberikan batasan nilai kisaran dari panjang gelombang yang akan dibandingkan sehingga nilai dari panjang gelombang yang berupa signal suara tersebut mempunyai nilai kisaran panjang yang sama. Daftar program untuk proses segmentasi dapat dilihat pada Tabel 3.



Gambar 4. Algoritma Pemilihan diphone 'ai' dan 'ay' pada suku kata yang mengandung diftong 'ai'.

Tabel 1.
Daftar program diphone1 dan diphone2.

| Diphone1 | Diphone2 |
|--|---|
| <pre> function popup2_Callback(hObject, eventdata, handles) angka2=get(hObject,'Value'); huruf2=get(hObject,'String'); switch huruf2{angka2}; case 'ai' handles.data2=wavread ('aiok.wav'); O=handles.data2; axes(handles.output2); plot(O); grid on case 'ay' handles.data2=wavread ('ayok.wav'); O=handles.data2; axes(handles.output2); plot(O); grid on end guidata(hObject,handles); </pre> | <pre> function popup3_Callback(hObject, eventdata, handles) angka3=get(hObject,'Value'); huruf3=get(hObject,'String'); switch huruf3{angka3}; case 'ai' handles.data3=wavread ('aiok.wav'); P=handles.data3; axes(handles.output3); plot(P); grid on case 'ay' handles.data3=wavread ('ayok.wav'); P=handles.data3; axes(handles.output3); plot(P) grid on end guidata(hObject,handles); </pre> |

Tabel 2.
Daftar program diphone1 dan diphone2.

```
function popup1_Callback(hObject, eventdata, handles)
angka1=get(hObject,'Value');
huruf1=get(hObject,'String');
switch huruf1{angka1};
    case 'pandai'
        handles.data1=wavread('pandai1.wav');
        Z=handles.data1;
        axes(handles.output1);
        plot(Z);
        grid on
    case 'ramai'
        handles.data1=wavread ('ramai1.wav');
        Z=handles.data1;
        axes(handles.output1);
        plot(Z);
        grid on
    case 'tunai'
        handles.data1=wavread ('tunai1.wav');
        Z=handles.data1;
        axes(handles.output1);
        plot(Z);
        grid on
    case 'tangkai'
        handles.data1=wavread ('tangkai1.wav');
        Z=handles.data1;
        axes(handles.output1);
        plot(Z);
        grid on
    case 'perisai'
        handles.data1=wavread ('perisai1.wav');
        Z=handles.data1;
        axes(handles.output1);
        plot(Z);
        grid on
    case 'lunglai'
        handles.data1=wavread ('lunglai1.wav');
        Z=handles.data1;
        axes(handles.output1);
        plot(Z);
        grid on
    case 'gapai'
        handles.data1=wavread ('gapai1.wav');
        Z=handles.data1;
        axes(handles.output1);
        plot(Z);
        grid on
end
guidata(hObject,handles);
```

Tabel 3.

Daftar program pemrosesan segmentasi

```
seg1=sig2(10:2100);
seg2=sig1(10:2100);
seg4=sig4(10:2100);
seg3=seg2-seg1;
seg5=seg2-seg4;
ns=16; nn=16;
Keterangan:
Sig2 merupakan input diphone 1
Sig4 merupakan input diphone 2
Sig1 merupakan input suku kata
```

Transformasi Fourier mempunyai persamaan seperti yang ditunjukkan Persamaan (1).

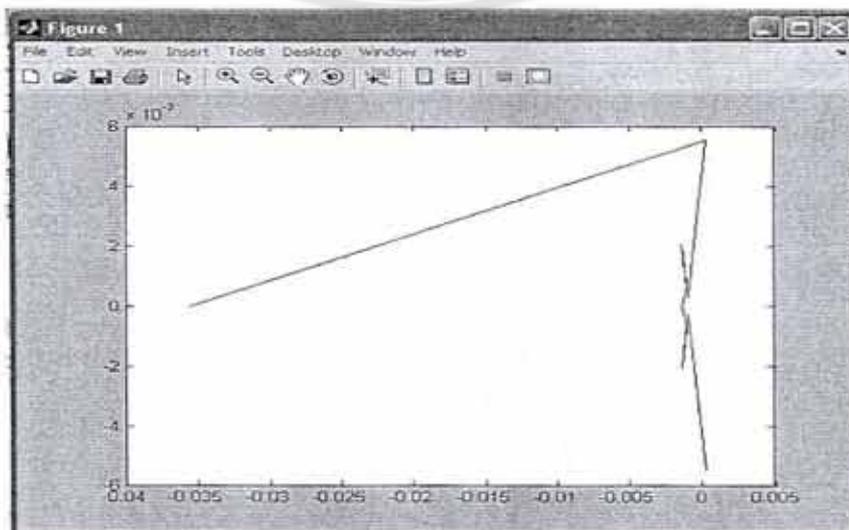
$$(1) \quad X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$$

Nilai kisaran yang ditentukan setelah proses segmentasi ditransformasi menggunakan transformasi Fourier yang menghasilkan dua buah nilai yaitu nilai nyata dan imajiner. Tabel 4 menunjukkan nilai dari diphone 'ai' yang telah ditransformasi Fourier, dengan panjang nilai yang diambil dari 0 sampai 10. Panjang gelombang yang digunakan adalah 0-20 untuk signal suara diphone 'ai,' tetapi pada signal suara diphone 'ai' yang sesungguhnya yang terdapat pada tampilan GUInya menggunakan panjang gelombang dari 0-2000. Gambar 5 merupakan grafik yang menjelaskan hasil transformasi Fourier dari diphone 'ai'.

Tabel 4.

Daftar program pemrosesan segmentasi

| | |
|--------------------------|--------------------------|
| -0.041748 | -0.0015963 -4.6372e-005i |
| -0.0083862 + 0.019186i | -0.0016075 - 0.0005212i |
| -0.0010204 + 0.0047425i | -0.0015741 - 0.0014059i |
| 0.00032622 + 0.0043376i | -0.0016346 - 0.0019624i |
| -0.0020852 + 0.0013625i | -0.0040283 - 0.0014648i |
| -0.0040283 + 0.0014648i | -0.0020852 - 0.0013625i |
| -0.0016346 + 0.0019624i | 0.00032622 - 0.0043376i |
| -0.0015741 + 0.0014059i | -0.0010204 - 0.0047425i |
| -0.0016075 + 0.0005212i | -0.0083862 - 0.019186i |
| -0.0015963 +4.6372e-005i | |
| -0.001709 | |



Gambar 5. Grafik 'ai' hasil transformasi Fourier

nilai absolut dari Tabel 4 dicari, sehingga dihasilkan nilai pada Tabel 5.

Tabel 5.
Harga mutlak

| |
|-----------|
| 0.041748 |
| 0.020939 |
| 0.0048511 |
| 0.0043498 |
| 0.0024908 |
| 0.0042864 |
| 0.002554 |
| 0.0021105 |
| 0.0016899 |
| 0.001597 |
| 0.001709 |
| 0.001597 |
| 0.0016899 |
| 0.0021105 |
| 0.002554 |
| 0.0042864 |
| 0.0024908 |
| 0.0043498 |
| 0.0048511 |
| 0.02093 |

Gambar 6 menunjukkan spektrum dari diphone 'ai' dimana panjang kisaran signal suara yang diambil dari 0 sampai 20. Karena pendeknya kisaran gelombang maka grafik spektrumnya pun terlihat sederhana. Nilai hasil inilah yang akan diestimasi menggunakan model AR ke dalam matriks ruang keadaan 16 x 16.

Proses transformasi Fourier dilakukan dengan menjalankan daftar program berikut:

```
ss1=sfft(seg1,2000,4);
ss3=sfft(seg3,2000,4);
ss4=sfft(seg4,2000,4);
ss5=sfft(seg5,2000,4);
```

Perkiraan Parameter Spektrum Diphone dan Derau Menggunakan Model AR

Inputan dari proses perkiraan ialah berupa signal suara dan signal derau dimana kedua signal tersebut telah ditransformasi fourier. Di bawah ini ialah persamaan model AR, dimana

input dari persamaan tersebut ialah nilai mutlak yang dihasilkan pada transformasi Fourier,

$$y(t) = \sum_{i=1}^n a_i y(t-i) + b v(t), \quad b \geq 0$$

$$x(t) = [y(t-n+1), y(t-n+2), \dots, y(t-1), y(t)]^T$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ a_n & a_{n-1} & a_{n-2} & \dots & a_1 \end{bmatrix}$$

$$x(t+1) = Ax(t) + b h v(t), \quad h = (0, 0, \dots, 0, 1)^T$$

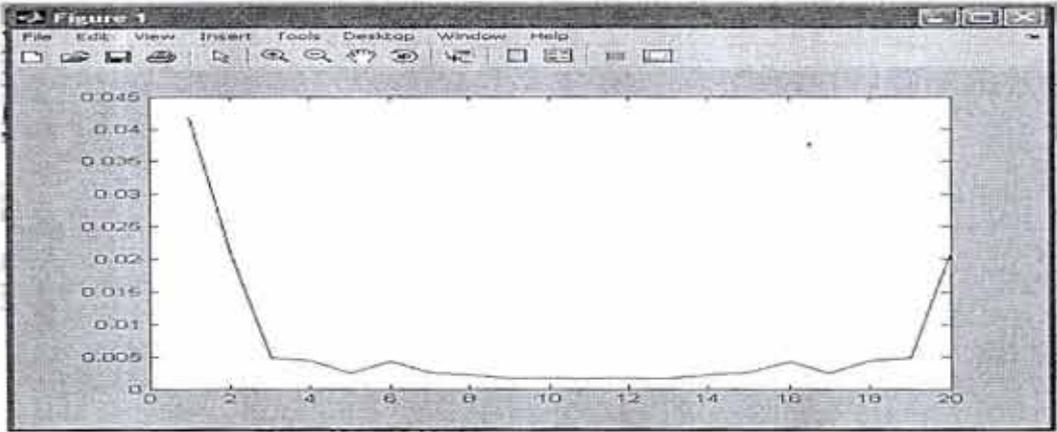
$$y(t) = h^T x(t)$$

Perkiraan parameter a dan b untuk model AR dari spektrum signal suara diphone 'ai' diperoleh dengan cara berikut::

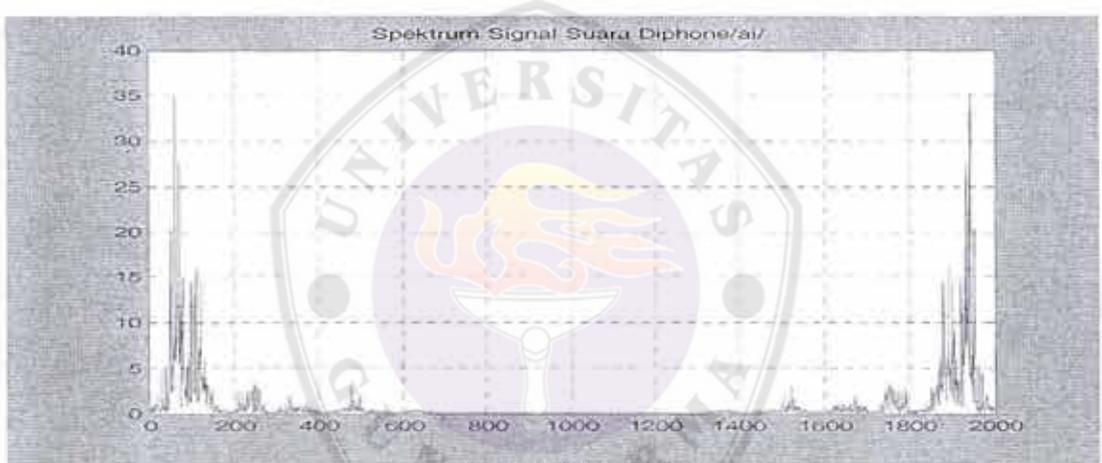
1. Untuk perkiraan spektrum dari signal suara (as dan bs) terdiri dari, $as = [3.1641 \quad -5.4225 \quad 7.3034 \quad -8.5322 \quad 8.697 \quad -7.8399 \quad 6.1535 \quad -4.0375 \quad 2.0517 \quad 0.84621 \quad 0.30227 \quad -0.023878 \quad -0.13742 \quad 0.071715 \quad 0.15531 \quad -0.15776]$, dan $bs = 0.0011705$ (Gambar 7).
2. Untuk perkiraan spektrum dari signal derau (an dan bn) terdiri dari, $an = [3.146 \quad -5.2163 \quad 6.6581 \quad -7.264 \quad 6.7681 \quad -5.3726 \quad 3.4311 \quad -1.4008 \quad -0.19445 \quad 0.81552 \quad -0.71878 \quad 0.49576 \quad -0.35938 \quad 0.14004 \quad 0.14474 \quad -0.15143]$, dan $bn = 0.001258$ (Gambar 8).

Matriks keadaan A yang diperoleh ditunjukkan Gambar 9, di mana $a1 = -0.15776$, $a2 = 0.15531$, $a3 = 0.071715$, $a4 = -0.13742$, $a5 = -0.023878$, $a6 = 0.30227$, $a7 = -0.84621$, $a8 = 2.0517$, $a9 = -4.0375$, $a10 = 6.1535$, $a11 = -7.8399$, $a12 = 8.697$, $a13 = -8.5322$, $a14 = 7.3034$, $a15 = -5.4225$, dan $a16 = 3.1641$.

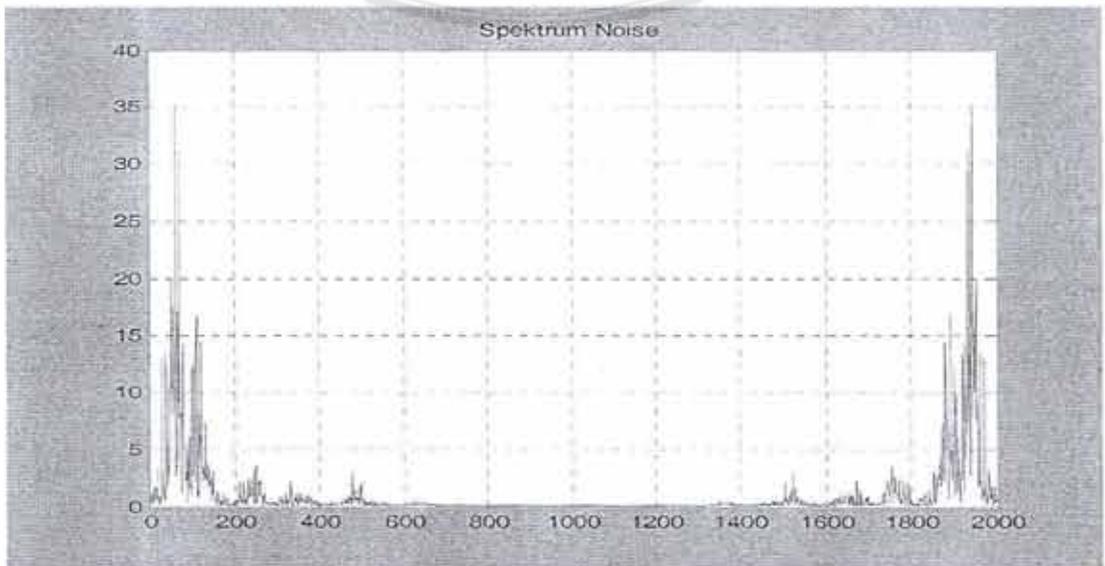
Matriks C (1x16) adalah $[0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1]$, dengan nilai $r1 = 0.0043817$. Matriks Q(16x16) dapat dilihat pada Gambar 10



Gambar 6. Gambar Spektrum diphone 'ai'



Gambar 7. Spektrum Diphone 'ai'



Gambar 8. Spektrum Derau


```
Q1(ns,ns)=bs1*bs1 ;
Q2=zeros(ns) ;
Q2(ns,ns)=bs2*bs2 ;
r1=sum(seg3.^2)/length(seg3) ;
r2=sum(seg5.^2)/length(seg5) ;
y1=kalmanf(seg2,A1,C,Q1,r1) ;
```

Penyaring Kalman

Ketiga matriks ruang keadaan yang telah terbentuk di atas akan dibandingkan dengan diphone yang dimasukkan. Hasil perbandingan karakteristik masing-masing diphone dengan matriks ruang keadaan yang telah terbentuk akan menghasilkan signal keluaran yaitu berupa Y1 dan Y2. Hasil perbandingan di atas akan mendapatkan nilai

kesalahan untuk masing-masing diphone. Diphone yang mempunyai nilai kesalahan terkecil diambil untuk suku kata yang dimasukkan. Daftar program pada proses penyaring Kalman, yaitu:

```
y1=kalmanf(seg2,A1,C,Q1,r1) ;
y2=kalmanf(seg2,A2,C,Q2,r2) ;
```

Output berupa grafik spektrum, grafik signal keluaran, dan nilai kesalahan dari masing-masing diphone. Daftar program untuk menampilkan spektrum diphone1 dan diphone2, signal keluaran, dan untuk menampilkan nilai kesalahan serta diphone yang terpilih dapat dilihat pada Tabel 6.

Tabel 6.
Daftar program spektrum, signal keluaran dan nilai kesalahan

| | Diphone1 | Diphone2 |
|-----------------|--|--|
| Spektrum | ss1=sfft(seg1,2000,4) ; axes(handles.axes4); plot(ss1); grid on | ss4=sfft(seg4,2000,4) ; axes(handles.axes8); plot(ss4); grid on |
| Signal keluaran | y1=kalmanf(seg2,A1,C,Q1,r1) ; axes(handles.axes9); plot(y1); grid on | y2=kalmanf(seg2,A2,C,Q2,r2) ; axes(handles.axes10); plot(y2); grid on |
| Nilai kesalahan | e1=seg1'-y1; e2=seg4'-y2; eavdiphone1=sum(abs(e1))/length(e1) set(handles.output1,'String',eavdiphone1); eavdiphone2=sum(abs(e2))/length(e2) set(handles.output2,'String',eavdiphone2); if eavdiphone1<eavdiphone2; pilih = 'diphone1' set(handles.output3,'String','Diphone 1'); else pilih = 'diphone2' set(handles.output3,'String','Diphone 2'); end | |

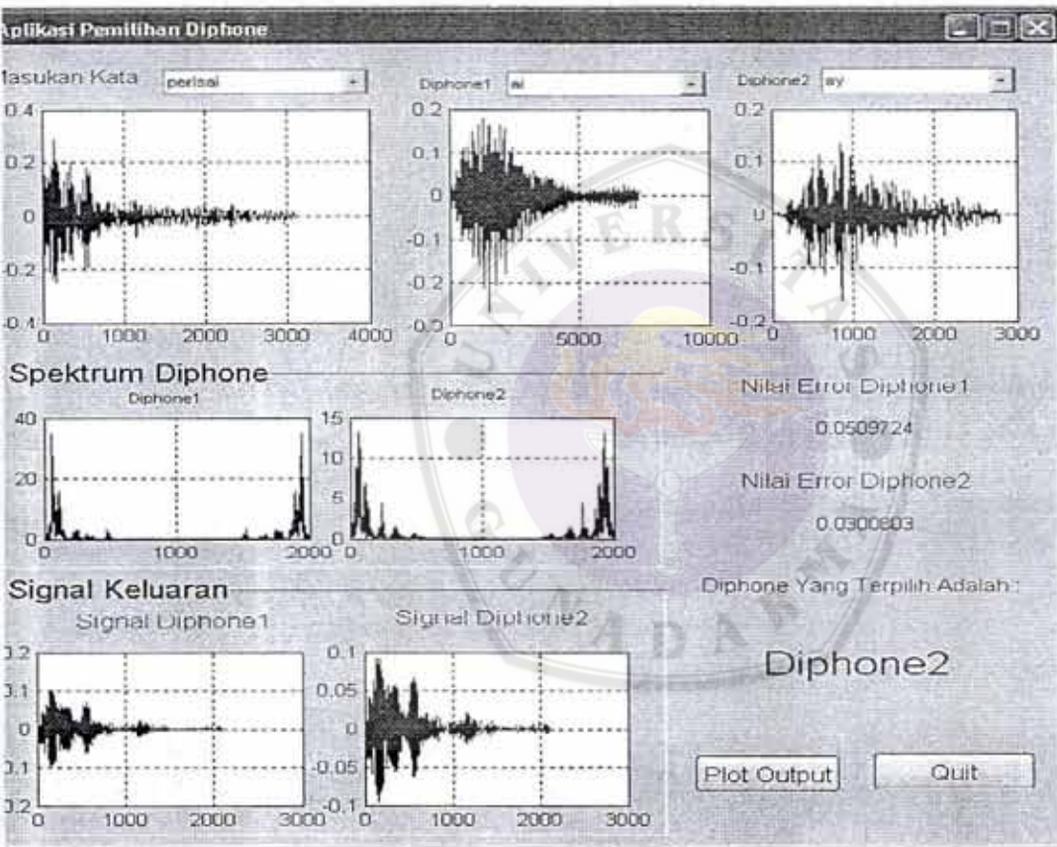
Tata Letak Tampilan GUI

Pada tampilan GUI, input yang dimasukkan berupa kata dan dua buah diphone. Inputan yang pertama ialah pada perintah 'Masukan Kata' dimana pada tool popup menu terdapat beberapa kata seperti 'pandai', 'tunai', 'tangkal', 'ramai', 'perisai', 'lunglai', 'gapai' setelah pemilihan kata pada tool popup menu pertama

selesai selanjutnya pemilihan diphone dimana terdiri dari diphone1 dan diphone2. Untuk tool popup menu diphone1 terdiri dari diftong 'ai' dan 'ay' begitu juga dengan tool popup menu diphone2 terdiri dari diftong 'ai dan 'ay' setelah semua inputan dimasukan maka proses pemilihan diphone untuk kata suatu kata akan diproses bila diklik tool push button 'Plot

put', setelah proses selesai maka akan dapat tampilan berupa grafik dan tulisan. Grafik terdiri dari tujuh buah grafik yang terdiri dari tiga buah grafik inputan dua buah grafik spektrum dari diphone yang kita pilih dan satu grafik hasil proses kalman filter. Untuk memilih diphone yang dipilih dan tulisan yang memberitahukan diphone mana yang dipilih pada kata yang dimasukkan. Push button 'Quit', berguna untuk keluar dari program tampilan GUI.

Pada Gambar 3.2b diphone1 dipilih diftong 'ai' dengan nilai kesalahan untuk kata 'perisai' sebesar 0.0509724 sedangkan diphone2 dipilih diftong 'ay' dengan nilai kesalahan untuk kata 'perisai' sebesar 0.0300803 maka diftong yang mempunyai nilai kesalahan terkecil lah yang diambil, pada contoh kata 'perisai' diftong yang diambil ialah diphone2 yaitu 'ay'. Hasil uji coba program dapat dilihat pada Tabel 7.



Gambar 9 Tampilan awal dari GUI ketika memilih kata 'perisai'

Tabel 7. Hasil Uji Coba Pemilihan Diftong

| Kata yang digunakan dalam pemilihan diphone | Nilai kesalahan diphone yang digunakan | | Diphone yang terpilih |
|---|--|------------|-----------------------|
| | Diphone ay | Diphone ai | |
| Pandai | 0.0266597 | 0.04622350 | ay |
| Ramai | 0.0237734 | 0.0444163 | ay |
| Tunai | 0.0242005 | 0.0444889 | ay |
| Tangkai | 0.0258793 | 0.046125 | ay |
| Perisai | 0.0300803 | 0.0509724 | ay |
| Lunglai | 0.0268108 | 0.0456808 | ay |
| Gapai | 0.0242768 | 0.0438531 | ay |

KESIMPULAN

Dari pembahasan di atas dapat disimpulkan bahwa simulator penyeleksian diphone untuk penggabungan suara pada pesintesa bahasa Indonesia dapat diperoleh hasil yang diinginkan. Diphone yang dipilih berdasarkan hasil percobaan adalah sesuai dengan kaidah bahasa Indonesia.

DAFTAR PUSTAKA

- A. Hunt and A. Black, "Unit Selection In A Concatenative Speech Synthesis Sistem Using A Large Speech Basis Data," *Proc. ICASSP*, pp. 373-376, 1996.
- E. Klabbbers and R. Veldhuis, "On The Reduction Of Concatenation Artefacts In Diphone Synthesis," *Proc. ICSLP98*, pp. 1983-1986, 1998.
- J. Wouters and M. Macon, "Perceptual Evaluation Of Distance Measures For Concatenative Speech Synthesis," *Proc. ICSLP98*, pp. 2747-2750, 1998.
- Y. Stylianou and Ann K. Syrdal, "Perceptual And Objective Detection Of Discontinuities In Concatenative Speech Synthesis," *Proc. ICASSP*, 2001.
- Robert E. Donovan, "A new distance measure for costing spectral discontinuities in concatenative speech synthesizers," *The 4th ISCA Tutorial and Research Workshop on Speech Synthesis*, 2001.
- J. Vepa, S. King, and P. Taylor, "Objective distance measures for spectral dscontinuities in concatenative speech synthesis," in *proc. ICLSP. Denver USA*, September 2002.
- J. Vepa, S. King, and P. Taylor, "New Objective distance measures for spectral and discontinuities in concatenative speech synthesis," in *proc. IEEE. workshop on Speech Synthesis*. Santa Monica, USA, September 2002.
- J. Frankel and S. King, "ASR-articulation speech recognition," in *Proc. Eurospeech, Aalborg, Denmark*, September 2000, pp. 599-602.
- G. Smith, J. de Fries, T. Robinson, and M. Niranjana, "Speech modeling using subspace and EM techniques," *Advances in Neural Information Processing Systems*, vol. 12, pp. 796-802, 1999.
- J. McKenna and S. Isard, "Tailoring Kalmaning towards Speakers Characterisation," in *Proc. Eurospeech, Budapest, Hungary, September 1999*, pp. 2793-2796.
- Z. Grahmani and G. Hinton, "Parameter Estimation for linear dynamical system," in *tech. rep. CRG-TR-96-2*. Dept of Computer Science. Univ of Toronto, 1996.
- Joe Frankel, "Linear dynamic models for automatic speech recognition," University of Edinburgh, April 2003.
- Jithendra vepa, Simon King, "Kalman-Filter based Join Cost For Unit-selection speech Synthesis," Centre for Technology Research, University of Edinburgh, 2003.
- Arry Akhmad Arman, "Proses pembentukan dan karakteristik signal ucapan", Teknik Elektro ITB, Juni 2003.
- Arry Akhmad Arman, "Perkembangan teknologi TTS Dari masa ke Masa", Teknik Elektro ITB, 2003.
- L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.