

ANALISA PENGGUNAAN METODE FILTER BANK PADA KONVERSI SPEECH TO TEXT LAYANAN MAILBOX

Reni Soelistijorini, Mike Yuliana, Rizki Nurhidayati
Prodi Teknik Telekomunikasi
Politeknik Elektronika Negeri Surabaya
Kampus PENS Keputih Sukolilo Surabaya 60111
Telp (+62)31-5947280, 5946114, Fax. (+62)31-5946114
Email: reni@eepis-its.edu, mieke@eepis-its.edu, rizki1403@yahoo.co.id

Abstrak

Layanan mailbox merupakan sebuah pesan yang direkam dalam bentuk suara yang kemudian disimpan pada mailbox penerima pesan. Pada penelitian ini, pesan yang ditinggalkan akan dikonversikan ke dalam media lain dengan memanfaatkan aplikasi IVR (Interactive Voice Response). File-file *.wav yang dihasilkan akan diolah menjadi file text dengan menggunakan metode Filter Bank. Data text yang dihasilkan akan dikirim menuju Mail Server, sehingga dapat diakses berdasarkan alamat Email yang dimiliki oleh user tujuan. Dari hasil pengujian dengan menggunakan metode Filter Bank, pada pengujian dua dan tiga data standar wanita memiliki prosentase keberhasilan lebih tinggi yaitu sebesar 58,3 % dan 46,665%. Sedangkan untuk satu data standar prosentase keberhasilan pria lebih tinggi yaitu sebesar 55,8%. Untuk keseluruhan pengujian, 2 data standar memiliki prosentase keberhasilan tertinggi yaitu sebesar 56,41 %.

Kata kunci : : mailbox, IVR, Filter Bank, data standar

Abstract

*Mailbox service is a message recorded in form of voice which is then stored in the message recipient mailbox. In this study, the messages received will be converted into the other media by utilizing IVR (Interactive Voice Response) application. The *.wav files produced will be processed into a text file by using a Filter Bank method. The textual data generated will be sent to the Mail Server, so that the data can be accessed by the destination users by using their email addresses. Based on the test results using the Filter Bank, for two and three standard data, women data have a higher percentage of success equal to 58.3% and 46.665%. As for one standard data, men data have higher percentage of success equal to 55.8%. For the entire test, two standard data have the highest percentage of success, that is 56.41%.*

Keywords : mailbox, IVR, Filter Bank, standard data

PENDAHULUAN

IVR adalah teknologi telepon yang memungkinkan komputer untuk mendeteksi suara dan penekanan tombol *keypad* telepon. Sistem IVR dapat merespon setiap penekanan tombol *keypad* telepon secara dinamis dengan menjalankan suara yang telah direkam sebelumnya, untuk mengarahkan penelepon. IVR dapat diimplementasikan pada layanan *mailbox*. Layanan ini banyak digunakan pada kehidupan sehari-hari, dimana kegunaannya adalah untuk meninggalkan pesan apabila panggilan tersebut tidak diangkat. Ketika penelepon melakukan panggilan pada sebuah nomor telepon maka pe-

manggil akan mendengar nada sambung. Setelah beberapa kali dering tidak diangkat, maka pelanggan akan terhubung ke server IVR yang meminta untuk meninggalkan pesan.

Layanan *Mailbox* ini memiliki beberapa kelemahan yaitu pesan yang ditinggalkan mengalami cacat atau data yang didapatkan tidak sempurna, sehingga banyak *user* yang akhirnya tidak menggunakan layanan ini dikarenakan layanan ini telah merugikan pihak pengguna dan menyebabkan pulsa terbuang sia-sia. Pada penelitian ini, pesan yang ditinggalkan akan dikonversikan ke dalam media lain dengan memanfaatkan ap-

likasi IVR (*Interactive Voice Response*). File- file *.wav yang dihasilkan oleh IVR akan diolah langsung ke dalam proses konversi *Speech to Text* dengan pemrograman C++. Data text yang dihasilkan akan dikirim menuju *Mail Server*, sehingga dapat diakses berdasarkan alamat *email* yang dimiliki oleh *user* tujuan.

Metode yang digunakan untuk konversi *Speech to Text* adalah *Filter Bank*. Metode ini menggunakan representasi sinyal konvolusi dalam melakukan *filter* terhadap sinyal. Setelah sinyal suara mempunyai nilai koefisien, maka nilai akan diolah untuk proses pencocokan dengan metode *Dynamic Time Warping*. *Dynamic Time Warping* ini bekerja dengan cara membandingkan dua buah data yaitu data *input* dan data standard yang digunakan sebagai *database*. Dimana selanjutnya dicari nilai yang terkecil dan dapat dinyatakan sebagai sinyal yang dimaksud. Dari nilai jarak yang terkecil ini dapat dilakukan konversi teks.

DASAR TEORI

Interactive Voice Response (IVR)

Interactive Voice Response (IVR) adalah teknologi teleponi dimana sebuah komputer dapat mendeteksi suara (*voice*) dan penekanan tombol pesawat telepon dengan menggunakan panggilan telepon normal. Sistem IVR dapat merespon setiap penekanan tombol telepon secara dinamis dengan menjalankan suara yang telah direkam sebelumnya, untuk mengarahkan penelepon atau *user* yang melakukan panggilan. Sistem IVR ini dapat berfungsi untuk mengontrol berbagai macam kebutuhan yang diinginkan dan fungsi umum dari IVR adalah untuk mengontrol *user* dalam jumlah besar.

IVR mempunyai fungsi yang cukup banyak diantaranya pengenalan DNIS (*Dialed Number Identification Service*), kemampuan konversi *text to speech*, memiliki kemampuan *speech recognition*, dapat melakukan *monitoring* dan sebagainya. Aplikasi yang da-

pat dibuat dari sistem IVR adalah membuat server layanan berbagai bidang yang membutuhkan jasa pelayanan (Yuliana, dkk, 2010); (Yuliana, 2008).

Layanan Mailbox

Mailbox merupakan fasilitas untuk mengalihkan panggilan ke kotak suara dalam kondisi tertentu, misalnya ketika *handphone* atau nomor yang dituju sedang tidak aktif, sedang melakukan panggilan, maupun tanpa kondisi artinya semua panggilan langsung dialihkan ke *mailbox*. Sistem akan menyimpan rekaman tersebut kemudian memberikan notifikasi kepada *user* bahwa yang bersangkutan menerima pesan suara serta instruksi untuk memutarinya. Sistem kerja dari *mailbox* yaitu data atau pesan yang telah direkam akan diubah ke dalam data digital kemudian disimpan dalam *mailbox storage (Harddisk)*, ketika penerima membuka *mailbox* nya maka pesan suara digital akan diubah kembali ke dalam data analog berupa suara dan akan langsung memutar pesan yang ada pada *mailbox storage* (Bellamy, 2005).

Speech Recognition

Speech recognition merupakan suatu proses identifikasi suara berdasarkan kata yang diucapkan dengan melakukan konversi sebuah sinyal akustik yang ditangkap oleh *audio device* (perangkat input suara). *Speech recognition* juga merupakan sistem yang digunakan untuk mengenali perintah kata dari suara manusia dan kemudian diterjemahkan menjadi suatu data yang dimengerti oleh komputer. Pada saat ini, sistem ini digunakan untuk menggantikan peranan input dari *keyboard* dan *mouse*.

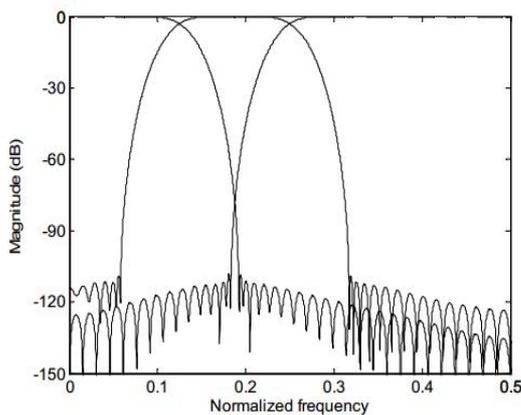
Keuntungan dari sistem ini adalah pada kecepatan dan kemudahan dalam penggunaannya. Parameter yang dibandingkan adalah tingkat penekanan suara yang kemudian akan dicocokkan dengan *template database* yang tersedia.

Filter Bank

Filter bank merupakan salah satu model pengenalan sinyal wicara yang dapat didefinisikan sebagai *array pass filter* yang memisahkan sinyal masukan menjadi beberapa komponen, masing-masing membawa satu subband frekuensi sinyal yang asli. *Filter bank* merupakan cara untuk dapat kembali kepada sinyal asli. *Filter bank* dapat digunakan untuk mengisolasi komponen frekuensi dalam sinyal jika perbedaan kecil pada frekuensi dianggap signifikan. (Ghosh. P.K, *et al* (2011); Jung, H (2004); Shannon, *et al* (2003).

Gambar 1 merupakan hasil dari proses filter bank 3 *channel*, dimana nilai pada sumbu x merupakan nilai dari *normalized frequency* dan nilai pada sumbu y merupakan nilai dari *magnitude* dalam satuan *db*. *Filter* digital yang digunakan adalah *FIR filter*. Rumus yang digunakan untuk *FIR filter* adalah :

$$y[n] = x[n]*b[0] + x[n-1]*b[1] + + x[n - M + 1]*b[M - 1] \dots\dots(1)$$



Gambar 1. *Filter bank* dengan 3 *channel* (*Magnitude response*)

DTW (Dynamic Time Warping)

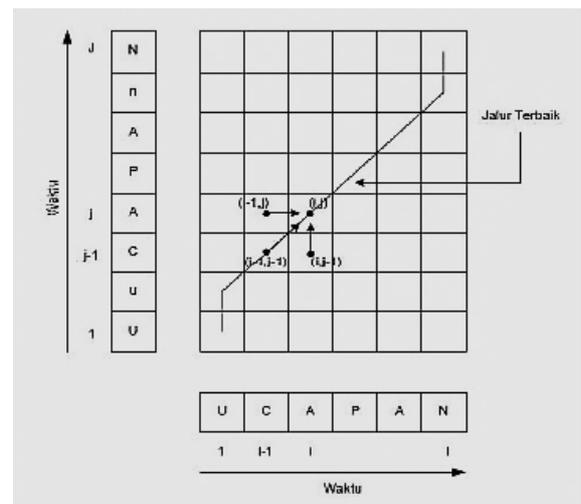
Metode *DTW (Dynamic Time Warping)* adalah metode yang digunakan untuk membandingkan suara input pembicara dengan

sampel suara yang ada di *database* dan menghasilkan keluaran data suara pembicara. Keunggulan *DTW* dari metode jarak yang lainnya adalah mampu menghitung jarak dari dua vektor data dengan panjang berbeda. Jarak diantara dua vektor dihitung dari jalur pembengkokan .

Dari beberapa teknik yang digunakan untuk menghitung *DTW*, salah satu yang paling handal adalah dengan metode pemrograman dinamis. Jarak *DTW* dapat dihitung dengan persamaan 2 dan persamaan 3 :

$$D(u, v) = y(m, n) \tag{2}$$

$$y(m, n) = d_{base}(u_i, v_j) + \min \begin{cases} y(i-1, j) \\ y(i-1, j-1) \\ y(i, j-1) \end{cases} \tag{3}$$



Gambar 2. Algoritma *DTW* untuk mengetahui hasil jarak terbaik “kata ucapan”

Gambar 2 menunjukkan bahwa *dynamic time warping* bekerja berdasarkan pencocokan pola, dengan jalan membandingkan jarak, yaitu jarak pada posisi diagonal bawah (*i-1,j-1*), samping kiri (*i-1,j*) dan bawah (*i,j-1*) jarak manakah yang paling kecil jarak itulah yang dipilih. Dari Gambar 2 terlihat bahwa jarak dari lokal yang ada (*d*) dapat dihitung. Apabila *d(i,j)* adalah jarak lokal

pada posisi (i,j) pada matrik, maka jarak $D(i,j)$ pada posisi (i,j) dengan kondisi awal (inisialisasi) $D(1,1) = d(1,1)$. Nilai jarak terakhir $D(I,J)$ pada posisi (I,J) pada matrik, menyatakan jarak pencocokan keseluruhan untuk pencocokan sebuah *template* referensi yang memiliki jarak pencocokan keseluruhan yang terkecil, atau nilai $D(I,J)$ terkecil bila dibandingkan nilai $D(I,J)$ dari pencocokan dengan *template* referensi lainnya.

Perancangan dan Pembuatan Sistem

Pada Penelitian ini dibuat sebuah sistem *mailbox* yang terintegrasi dengan email, adapun alur kerja dari sistem yang akan dibuat adalah sebagai berikut :

1. Pada sisi *user*, terdapat layanan *Mailbox*. Apabila *user* melakukan panggilan menuju ke *user* lain namun *user* yang dituju tidak merespon panggilan, *user* yang melakukan panggilan dapat meninggalkan pesan pada layanan *Mailbox* tersebut.
2. Data yang diperoleh dari layanan *Mailbox* akan disimpan menjadi file berupa file *wav* di dalam *Mailbox Server*.
3. Setelah itu file diproses ke dalam proses *Speech to Text*, agar file dapat diakses atau tersimpan ke dalam visual dalam format *text*.
4. Setelah proses konversi dari *Speech* menuju *Text*, maka file *text* tersebut dikirim menuju layanan *Mail Server* agar dapat dikirimkan ke alamat *email* dari *user* yang dituju dan dapat diakses kapanpun oleh *user* penerima. Sehingga pesan yang dikirim dapat diterima secara utuh dan tidak mengalami kerusakan data.

Perancangan Sistem Layanan Mailbox

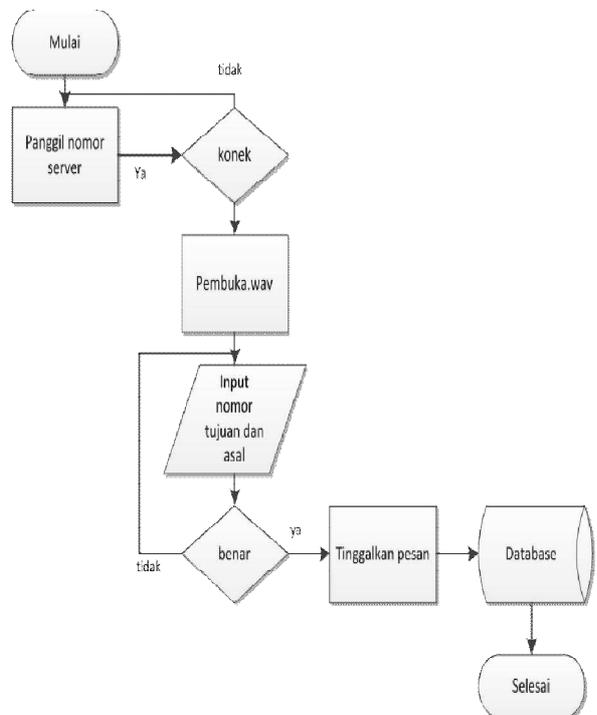
Sistem layanan mailbox pada penelitian ini berfungsi untuk mendapatkan sinyal wicara yang akan diolah pada proses *speech to text* format *file* yang diperoleh dari proses *mailbox* ini adalah file **.wav*. Untuk kata-kata yang diucapkan untuk penelitian ini

juga sudah ditentukan diawal, sehingga kata-kata yang direkam juga disesuaikan dengan daftar kata yang telah terdaftar untuk penelitian ini. Daftar kata yang digunakan untuk penelitian ini sesuai dengan Tabel 1.

Tabel 1. Daftar kata file **.wav* yang digunakan pada penelitian

Daftar kata
bertemu
halo
jumpa
kembali
pesan
telfon

Dari daftar kata pada Tabel 1 tersebut dilakukan proses perekaman pada beberapa *user*, dimana *user* yang digunakan di dalam penelitian ini berjumlah 4 (empat) orang (2 (dua) *user* pria dan 2 (dua) *user* wanita). Setiap *user* pada proses perekaman ini melakukan pengucapan daftar kata sejumlah 10 (sepuluh) kali pengucapan untuk pengujian dari penelitian ini.



Gambar 3. Flowchart mailbox server

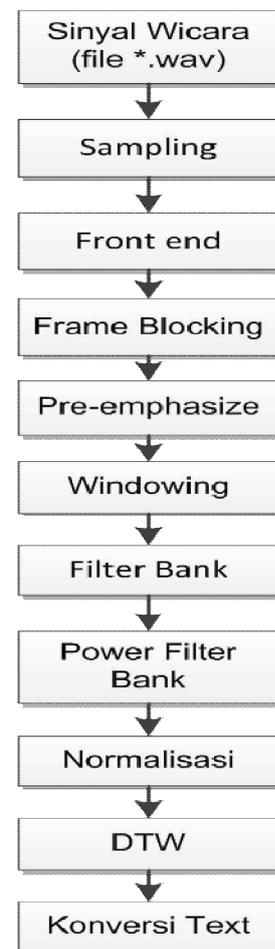
Gambar 3 merupakan tahapan-tahapan yang dilakukan pada sistem *mailbox*. Sistem *mailbox* ini harus memiliki satu nomor ekstensi yang digunakan sebagai nomor server. Sistem *mailbox* dapat dijalankan hingga menghasilkan file suara atau pesan yang telah direkam dengan cara melakukan panggilan menuju nomor server, setelah koneksi berhasil maka dari sistem IVR akan langsung memainkan file wav berupa ucapan pembuka sebagai tanda masuk ke dalam sistem *mailbox*.

Perancangan Sistem Kerja Proses *Speech to Text*

Gambar 4 menunjukkan tahapan-tahapan yang dilakukan dalam proses *Speech to Text* adalah file wicara yang diperoleh dari layanan *Mailbox* berupa file wav digunakan sebagai *input* proses *speech to text*. Dari *input* sinyal wicara yang ada inilah dilakukan proses pemotongan sinyal yang tidak mengandung sinyal informasi. Selanjutnya file wicara diolah dengan proses *frame blocking*. Proses yang dilakukan ini biasa disebut dengan proses *framing*. Dari poses *framing*, dilanjutkan dengan proses *pre-emphasize* dan *windowing*. Setelah proses *windowing* selesai langsung diproses ke dalam metode yaitu *filter bank*. Pada proses ini terdapat dua tahapan yang akan dilakukan yaitu proses mencari nilai koefisien *filter bank* dan konvolusi dari *filter bank* itu sendiri. Nilai koefisien *filter bank* diperoleh dari nilai *band pass filter* yang digunakan. Sedangkan untuk nilai konvolusi diperoleh dari pemanfaatan rumus FIR *filter*.

Dari hasil proses konvolusi ini, maka diperoleh nilai *filter bank* yang akan digunakan untuk proses pemadanan dengan DTW. Namun sebelum masuk ke dalam proses DTW perlu dilakukan pengambilan *power* dari nilai konvolusi *filter bank*. Setelah nilai *power* diperoleh, dilakukan proses normalisasi, proses ini digunakan untuk menyesuaikan amplitudo dari sinyal wicara agar berada diantara range 0 hingga 1. Apabila nilai nor-

malisasi sudah terpenuhi maka dapat dilakukan proses pencocokan data yang diuji dengan data yang terdapat pada *database*. Apabila data cocok dengan salah satu data yang ada pada *database* maka dapat dilakukan konversi ke dalam teks. Dan tampilan untuk hasil teks setelah dilakukan proses DTW ini adalah pada notepad atau pada file dengan ekstensi .txt.



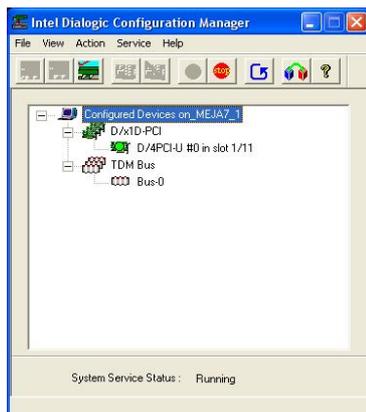
Gambar 4. Blok diagram sistem kerja proses *Speech to Text*

UJI COBA DAN ANALISA Implementasi Model

Pada penelitian ini, untuk sistem layanan mailbox digunakan perangkat lunak yang disebut sebagai sistem *Interactive Voice Response (IVR)* yaitu *Intel Dialogic System Software* Seperti yang terlihat pada Gambar

5. Setelah *software* dikonfigurasi, maka baru bisa dilakukan proses perekaman suara, dengan membuka bagian *Sample Dialogic Voice Application* pada *software Intel Dialogic System*. Sebelum melakukan perekaman suara, perlu dilakukan konfigurasi format perekaman yang akan dilakukan. Konfigurasi ini dilakukan pada pilihan menu *options-Record WAVE format*, dilanjutkan dengan memilih format yang digunakan. Dalam penelitian ini, format yang digunakan adalah *Linear PCM 8KHz*. Format ini dapat diinisialisasi menjadi parameter sebagai berikut :

1. Sampling rate : 8.000 Hz
2. Jumlah Bit / Resolusi Data : 8 bit (LIN8)



Gambar 5. Tampilan Intel Dialogic Configuration Manager

Pengujian Layanan Mailbox Server

Hasil pengujian dari sistem layanan *mailbox* digambarkan melalui hasil dari proses melakukan panggilan hingga perekaman pesan dan melihat hasil *update* dari *database* yang digunakan. Gambar 6 merupakan tampilan dari *database* yang telah dirancang untuk sistem layanan *mailbox*. Untuk awal sebelum melakukan proses perekaman atau menjalankan program *mailbox* perlu dilakukan pengisian terlebih dahulu pada *database* untuk nomor asal dan tujuan. Sehingga apabila program dijalankan dan ada kondisi yang digunakan, maka tabel akan langsung

melakukan proses *update* untuk kolom “pesan” dan “filename”.

id	asal	tujuan	pesan	filename
1	813	807	0	
2	814	807	0	

Gambar 6. Tampilan database sebelum proses rekam



Gambar 7. Tampilan hasil program mailbox

Gambar 7 menjelaskan tampilan hasil dari program *mailbox* yang dijalankan. Pada tampilan ini terlihat proses yang dijalankan, yaitu proses melakukan panggilan menuju nomor server, melakukan input nomor yang dituju, melakukan input nomor asal dan proses perekaman pesan. Sesuai dengan Gambar 7 ini terlihat proses panggilan yang dilakukan oleh user “814” dan ingin meninggalkan sebuah pesan yang ditujukan kepada user dengan ekstensi “807”.

id	asal	tujuan	pesan	filename
1	813	807	0	
2	814	807	1	pesan1a.WAV

Gambar 8. Tampilan database setelah program mailbox dijalankan

Tampilan pada Gambar 8 merupakan hasil *database* yang telah di *update*, isi dari “filename” adalah **pesan1a.wav** (merupakan hasil proses perekaman dari salah satu kata yang ada pada daftar kata Tabel 1). Dan isi dari kolom “pesan” untuk “id=2” terbarukan

menjadi “1”, hal ini menyatakan bahwa data untuk kondisi “asal=814” dan “tujuan = 807” sudah terisi.

Pengujian Dynamic Time Wrapping

Pengujian ini merupakan pengujian terakhir dari sistem pengolahan sinyal suara yaitu membandingkan sinyal hasil *filter bank* yang telah dinormalisasi antara data *input* dan data standar. Data yang digunakan di dalam pengujian *dynamic time warping* ini adalah 6 (enam) buah kata yang dilakukan sebanyak 10 (sepuluh) kali pengujian untuk masing-masing kata. Dalam hal ini, data yang diuji terdapat 10 (sepuluh) data dengan perbandingan 1 (satu) data standar, 2 (dua) data standar dan 3 (tiga) data standar. Dengan membandingkan jarak antar sinyal *input* dan sinyal standar, berikut ini adalah data jarak sinyal *input* dengan sinyal standar sesuai dengan jumlah data standar yang digunakan sebagai data pembandingan dengan data yang diuji di dalam program *speech to text*.

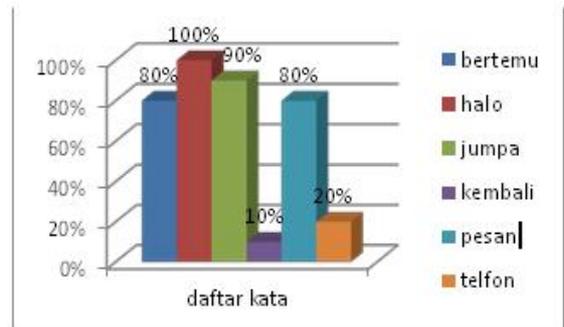
Tabel 2. Hasil proses DTW kata “pesan” untuk satu data standar

Pengujian ke-	Hasil Output Teks Kata "pesan"			
	pria 1	pria 2	wanita 1	wanita 2
1	pesan	pesan	bertemu	pesan
2	jumpa	pesan	pesan	pesan
3	pesan	pesan	pesan	pesan
4	pesan	pesan	pesan	pesan
5	pesan	pesan	kembali	pesan
6	pesan	pesan	pesan	pesan
7	pesan	hubungi	pesan	pesan
8	pesan	pesan	bertemu	pesan
9	pesan	pesan	kembali	pesan
10	bertemu	pesan	pesan	pesan
prosentase	80%	90%	60%	100%

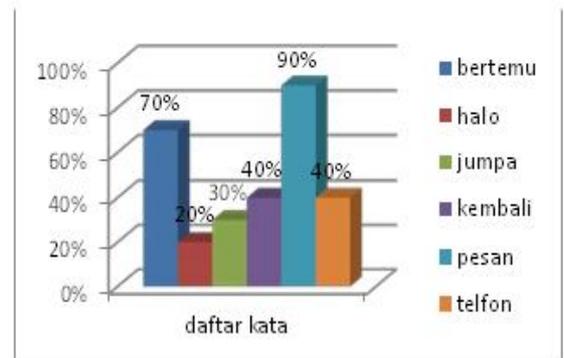
Sumber : Data Olahan (2014)

Tabel 2 menunjukkan contoh hasil pengujian dari pengujian proses DTW untuk kata “pesan”. Dari tabel tersebut dapat dilihat perbandingan antara nilai prosentase keberhasilan dengan suara wanita dan pria untuk kata “pesan”. Hasil prosentase ter-

besar untuk pendeteksian kata “pesan” pada penelitian ini adalah *user* wanita 2 dengan prosentase sebesar 100%. Sedangkan hasil prosentase rata-rata keseluruhan *user* yang didapatkan untuk kata “pesan” sebesar 82,5%.



Gambar 9. Prosentase keberhasilan *user* “pria 1” untuk semua kata dengan pengujian satu data standar



Gambar 10. Prosentase keberhasilan *user* “pria 2” untuk semua kata dengan pengujian satu data standar

Gambar 9 dan 10 merupakan contoh diagram batang dari prosentase keberhasilan *user* pria 1. Pengenalan terhadap kata yang tinggi adalah sebesar 100% dengan kata yang terdeteksi adalah “halo”. Sedangkan pada *user* pria 2 pengenalan terhadap kata yang tinggi adalah sebesar 90% dengan kata yang terdeteksi adalah “pesan”. Dan hasil pengenalan kata yang terendah dari “pria 1” ini adalah kata “kembali” dengan nilai prosentase yang sangat kecil yaitu 10%, sedangkan pengenalan kata yang terendah

dari “pria 2” ini adalah kata “halo” dengan nilai prosentase 20%.

Tabel 3. Prosentase Keberhasilan User untuk Satu Data Standar

Kata	Prosentase keberhasilan user (%)			
	Pria 1	Pria 2	Wanita 1	Wanita 2
bertemu	80	70	30	50
halo	100	20	100	50
jumpa	90	30	60	100
kembali	10	40	30	70
pesan	80	90	40	100
telfon	20	40	0	10
Rata-rata	63.3	48.3	43.3	63.3

Sumber : Data olahan (2014)

Tabel 4. Prosentase Keberhasilan User untuk Dua Data Standar

Kata	Prosentase keberhasilan user (%)			
	Pria 1	Pria 2	Wanita 1	Wanita 2
bertemu	60	100	70	100
halo	50	20	40	20
jumpa	90	30	100	30
kembali	10	50	50	80
pesan	100	100	70	100
telfon	30	20	20	20
Rata-rata	56.7	53.33	58.3	58.3

Sumber : Data olahan (2014)

Tabel 5. Prosentase Keberhasilan User untuk Tiga Data Standar

Kata	Prosentase keberhasilan user (%)			
	Pria 1	Pria 2	Wanita 1	Wanita 2
bertemu	60	40	60	50
halo	60	20	50	30
jumpa	70	20	20	80
kembali	10	40	20	60
pesan	70	70	60	90
telfon	10	10	30	10
Rata-rata	46.67	33.3	40	53.33

Sumber : Data olahan (2014)

Tabel 3-5 menunjukkan prosentase keberhasilan *user* dalam mengakses sistem. Dari tabel terlihat bahwa untuk dua dan tiga data standar prosentase keberhasilan wanita lebih tinggi yaitu sebesar 58,3 % dan 46,665%. Sedangkan untuk satu data standar prosentase keberhasilan pria lebih tinggi yaitu sebesar 55,8%. Hal ini dikarenakan,

penggunaan kata yang berbeda-beda meskipun kata yang diucapkan sama. Sehingga menyebabkan terjadinya perbedaan dalam pembentukan koefisien filternya dan hasil yang didapat juga menjadi bervariasi ada yang menghasilkan data jarak yang sesuai dan ada yang tidak sesuai. Nilai jarak terkecil ini dipengaruhi oleh pemilihan file sinyal standar yang digunakan. Apabila data standar yang digunakan kurang tepat maka hasil pencocokan data antara data standar dan data *input* akan menghasilkan prosentase yang rendah.

Tabel 6. Hasil prosentase keseluruhan/kata untuk tiap data standar

Kata	rata-rata prosentase keseluruhan/kata		
	1 data standar	2 data standar	3 data standar
bertemu	57.5%	82,5%	52.5%
halo	67.5%	32,5%	40%
jumpa	70.0%	63%	47.5%
kembali	37.5%	47,5%	32.5%
pesan	82.5%	93%	72.5%
telfon	17.5%	20%	15%

Sumber : Data olahan (2014)

Dari hasil pada Tabel 6 terlihat perbandingan nilai rata-rata prosentase keseluruhan untuk tiap kata pada masing-masing pengujian dengan jumlah data standar yang digunakan di dalam penelitian ini. Pada Tabel 6, dapat diambil kesimpulan bahwa hasil terbaik untuk hasil pencocokan DTW adalah menggunakan 2 (dua) file data standar untuk masing-masing kata yang digunakan. Meskipun pada kata “halo” dan kata “jumpa” masih lebih kecil dibandingkan dengan hasil prosentase untuk data standar yang lain. Tapi untuk keseluruhan pengujian hasil masih lebih baik untuk pengujian menggunakan 2 data standar yaitu sebesar 56.41 % dibandingkan dengan yang lain. Hal ini disebabkan karena 2 data standar ini terdiri dari 1 (satu) file dengan *user* pria dan 1 (satu) file dengan *user* wanita. Penentuan data standar ini sangat berpengaruh karena

user yang diuji pada penelitian ini adalah *independent speaker*, sehingga diperlukan satu untuk masing-masing *gender*.

KESIMPULAN

Dari hasil pengujian dan analisa dapat disimpulkan bahwa :

1. Rata-rata prosentase keberhasilan tertinggi adalah pada kata pesan untuk semua data standar yaitu 82,5% (satu data standar), 93% (dua data standar), 72,5% (tiga data standar).
2. Pada dua dan tiga data standar prosentase keberhasilan wanita lebih tinggi yaitu sebesar 58,3 % dan 46,665%. Sedangkan untuk satu data standar prosentase keberhasilan pria lebih tinggi yaitu sebesar 55,8%.
3. Untuk keseluruhan pengujian, 2 data standar memiliki prosentase keberhasilan tertinggi yaitu sebesar 56.41 %

DAFTAR PUSTAKA

Bellamy, J (2005) *Digital Telephony*, John Wiley & Sons.

Ghosh, P.K., Goldstein, M.L, and Narayanan, S.S (2011) Processing Speech Signal using Auditory-like Filterbank Provides Least Uncertainty about Articulatory Gestures, *The Journal of the Acoustical Society of America*, Vol. 129, No. 6, pp. 4014-4022.

Jung, H.Y (2004) Filtering of Filter-Bank Energies for Robust Speech Recognition, *ETRI Journal*, Vol. 26, No. 3, pp. 273-276.

Shannon, B.J., Kuldip K., Paliwal (2003) A Comparative Study of Filter Bank Spacing for Speech Recognition, *Micro-electronic Engineering Research Conference*, pp. 1-3.

Yuliana, M (2008) *Praktikum Sistem Voice Response (IVR)*, Modul dasar Jaringan Telephony, Politeknik Elektronika Negeri Surabaya.

Yuliana, M., Huda, M., dan Kristalina, P (2010) Analisa Kualitas Sinyal Suara pada Layanan Mailbox Berbasis Teknologi Interactive Voice Response (IVR), Politeknik Elektronika Negeri Surabaya.