

Analisa Kualitas Sinyal Suara pada Layanan MailBox Berbasis Teknologi Interactive Voice Response(IVR)

Mike Yuliana, Miftahul Huda, Prima Kristalina
Politeknik Elektronika Negeri Surabaya-Institut Teknologi Sepuluh Nopember
Kampus ITS Keputih Sukolilo Surabaya 60111, Indonesia
Tel:+62(31)5947280;Fax:+62(31)5946114
E-mail: mieke@eepis-its.edu

Abstrak

Salah satu Aplikasi IVR yang diimplementasikan pada layanan mailbox memiliki kegunaan untuk meninggalkan pesan jika penerima telepon tidak ada di rumah. Kendala yang dihadapi adalah kurang jernihnya pesan yang ditinggalkan, sehingga aplikasi IVR tersebut kurang memuaskan pengguna.

Pada penelitian ini akan dianalisa kualitas sinyal suara dengan menggunakan metode ACR(Absolute Category Rating) dan membandingkannya dengan SNR(Signal to Noise Ratio). Hasil analisa digunakan untuk menentukan format perekaman berbasis metode *companding* yang menghasilkan kualitas sinyal suara terbaik pada teknologi IVR yang diimplementasikan pada layanan mailbox.

Dari hasil pengujian diketahui bahwa dengan menggunakan metode ACR dan SNR kualitas sinyal suara terbaik didapatkan dengan menggunakan format perekaman berbasis PCM 8 KHz dengan nilai MOS sebesar 3.98 dan SNR sebesar 51.89 dB, sedangkan *sampling rate* yang memberikan kualitas sinyal suara terbaik pada seluruh format perekaman yang ada adalah 8 KHz.

Kata kunci: ACR, IVR, mailbox, MOS, PSD.

1. Pendahuluan

Saat ini layanan IVR banyak digunakan untuk berbagai sistem layanan seperti layanan tagihan PLN, tagihan telepon, informasi nilai sampai cerita anak-anak. Teknologi ini memanfaatkan database suara yang disimpan di dalam komputer. Database ini dipanggil melalui telepon, selanjutnya dimainkan sesuai dengan pilihan yang di *dial* oleh pelanggan melalui keypad telepon[1][2][4].

Dialogic card D/4 PCIU merupakan suatu perangkat yang digunakan sebagai *interface* antara sistem telepon dan komputer. Dengan menggunakan card tersebut bisa dibuat beberapa aplikasi yang antara lain meliputi *networked voice messaging, automated attendant, serta interactive voice response*. Untuk mendukung aplikasi IVR(Interactive Voice Response), terdapat fungsi untuk

memainkan dan merekam *file* suara. Dimana beberapa format yang didukung oleh Dialogic card D/4 PCIU, diantaranya ADPCM, PCM,ALAW dan MULAW. Pada penelitian ini akan dilakukan analisa kualitas sinyal suara hasil perekaman dengan berbagai format untuk mendapatkan format perekaman suara terbaik.

2. Penelitian Terdahulu

Mike Yuliana[3] membuat Aplikasi Informasi Nilai Indeks Prestasi Mahasiswa PENS ITS Berbasis *Computer Telephony Integration*(CTI), dimana pada penelitian ini dibuat suatu layanan informasi nilai indeks prestasi mahasiswa PENS-ITS. Prima Kristalina[5] membuat Analisa Tingkat Intelligibilitas Suara pada Layanan *Interactive Voice Response* dengan Perekaman Berbasis Metode *Companding* PCM dan ADPCM, dimana pada penelitian ini tingkat intelligibilitas suara diukur dengan menggunakan metode ACR sehingga didapatkan metode *companding* yang menghasilkan rekaman dengan tingkat kejernihan tertinggi.

Penelitian ini merupakan pengembangan dari 2 penelitian sebelumnya, dimana pada penelitian ini akan dilakukan analisa kualitas sinyal suara pada teknologi IVR tidak hanya pada metode *companding* PCM dan ADPCM namun juga ALAW dan MULAW, selain itu tidak hanya digunakan cara subyektif(ACR) namun juga dengan cara obyektif(SNR).

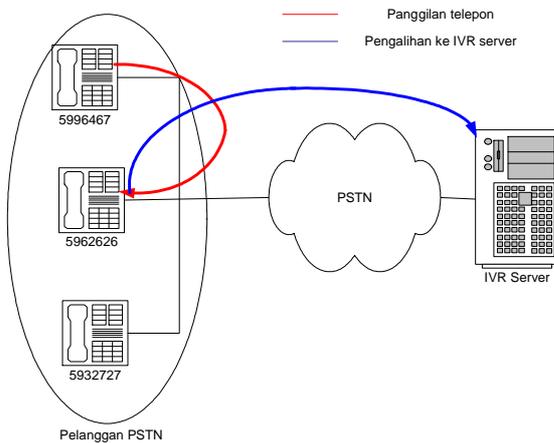
3. Dasar Teori Sistem

3.1 Teknologi IVR(Interactive Voice Response)

IVR adalah teknologi telepon yang memungkinkan komputer untuk mendeteksi suara dan penekanan tombol keypad telepon. Sistem IVR dapat merespon setiap penekanan tombol *keypad telephone* secara dinamis dengan menjalankan suara yang telah direkam sebelumnya, untuk mengarahkan penelepon.

Pada penelitian ini, teknologi IVR yang dibuat pada Gambar 1 diimplementasikan pada layanan mailbox. Layanan ini sangat *familiar* dengan kehidupan sehari-hari, dimana kegunaannya adalah untuk meninggalkan pesan jika panggilan tidak diangkat. Ketika penelepon melakukan

panggilan pada sebuah nomor telepon maka pemanggil akan mendengar nada sambung. Setelah beberapa kali dering tidak diangkat, maka pelanggan akan terhubung ke IVR server yang meminta untuk meninggalkan pesan



Gambar 1. Layanan Mailbox Berbasis IVR

3.2 Dialogic Card D/4 PCIU

Dialogic D/4 PCIU pada Gambar 2 merupakan suatu perangkat yang digunakan untuk mengolah suara dimana arsitekturnya mengkombinasikan kemampuan pengolahan sinyal dan penggunaan transfer data dengan menggunakan prosesor yang cepat dan menyediakan memori yang lebih besar. Arsitektur perangkat ini berfungsi untuk menangani kejadian – kejadian *real-time*, mengatur aliran data ke PC yang memerlukan respon waktu yang cepat, melayani permintaan pengolahan dari suatu PC, dan mengolah DTMF dan sinyal telepon dan menjalankan pengolahan sinyal pada *incoming call*. Beberapa aplikasi yang dapat dibangun Dialogic D/4 PCIU antara lain:

- *Automated attendant*
- *Interactive Voice Response*



Gambar 2. Dialogic Card D/4 PCIU

3.3 Perekaman Suara Berbasis Metode Companding

Perekaman suara yang akan dianalisa pada penelitian ini, menggunakan fungsi perekaman suara yang telah tersedia pada *driver Dialogic card*

```
void record(int chdev, char fname[10])
{
    int fd;
    /* file descriptor for file to be played */
    DX_IOTT iott; /* I/O transfer table */
    DV_TPT tpt; /* termination parameter table */
    DX_XPB xpb; /* I/O transfer parameter block */

    tpt.tp_type = IO_EOT;
    tpt.tp_termno = DX_MAXDTMF;
    tpt.tp_length = 1;
    tpt.tp_flags = TF_MAXDTMF;

    if((fd = dx_fileopen(fname, O_RDWR, O_BINARY)) == -1)
    {
        printf("File open error\n");
        exit(2);
    }

    /* clear digit buffer */
    dx_clrdigbuf(chdev);

    /* Set up DX_IOTT */
    iott.io_handle = fd;
    iott.io_bufp = 0;
    iott.io_offset = 0;
    iott.io_length = -1;
    iott.io_type = IO_DEV | IO_EOT;

    xpb.wFileFormat = FILE_FORMAT_WAVE;
    xpb.wDataFormat = DATA_FORMAT_PCM;
    xpb.nSamplesPerSec = DRT_11KHZ;
    xpb.wBitsPerSample = 11;

    if (dx_reciottdata(chdev, &iott, &tpt, &xpb, PM_TONE|EV_SYNC) == -1)
    {
        printf("Error recording file - %s\n",
            ATDV_ERRMSGP(chdev));
        exit(4);
    }
}
```

Disesuaikan dengan yang diinginkan

Gambar 3. Listing Program Perekaman File Suara

wFileFormat	Specifies one of the following audio file formats. Note that this field is ignored by the convenience functions <code>dx_recwav()</code> , <code>dx_recvox()</code> , and <code>dx_playvox()</code> . FILE_FORMAT_VOX Dialogic VOX file format FILE_FORMAT_WAVE Microsoft WAVE file format
wDataFormat	Specifies one of the following data formats: DATA_FORMAT_DIALOGIC_ADPCM 4-bit OKI ADPCM (Dialogic registered) DATA_FORMAT_MULAW 8-bit mu-law PCM DATA_FORMAT_ALAW 8-bit a-law PCM DATA_FORMAT_PCM 8-bit Linear PCM DATA_FORMAT_G726 G.726 bit-exact coder DATA_FORMAT_GSM610_MICROSOFT GSM 6.10 WAVE full-rate coder (Microsoft Windows compatible format)* DATA_FORMAT_GSM610_TIPHON GSM 6.10 WAVE full-rate coder (TIPHON format)**
nSamplesPerSec	Specifies one of the following sampling rates: DRT_6KHZ 6 KHz sampling rate. DRT_8KHZ 8 KHz sampling rate. DRT_11KHZ 11 KHz sampling rate. Note: 11KHz OKI ADPCM is not supported.
wBitsPerSample	Specifies the number of bits per sample. This field must be set to 8 for MULAW, ALAW, and PCM. Set the field to 4 for ADPCM. For G.726 and GSM, refer to the following information.

Gambar 4. Format-Format Jenis Perekaman Berbasis Metode *Companding*

Pada bagian ini, jenis-jenis perekaman yang akan ditentukan diubah-ubah sesuai dengan format data yang ada pada Gambar 4.

3.4 ACR(Absolute Category Rating)

Pada Metode ACR, responden mendengar sample suara terkode dan menilai kualitas suaranya pada suatu skala *absolute*. Hasil dari metode ACR ini berupa *Mean Opinion Score*(MOS). Dalam sistem *telephony*, MOS adalah faktor penting dalam menentukan *Quality Of Service* (QoS) dari suatu layanan suara. Tabel 1 memperlihatkan skala absolute pada metode ACR.

Tabel 1. Skala Absolut pada Metode ACR

SKALA ABSOLUT	KUALITAS	KETERANGAN
5	<i>Excellent</i>	Sangat jelas dan sangat jernih
4	<i>Good</i>	Jelas dan jernih
3	<i>Fair</i>	Cukup jelas dan cukup jernih
2	<i>Poor</i>	Tidak jelas dan tidak jernih
1	<i>Bad</i>	Sangat tidak jelas dan sangat tidak jernih

Pengujian dilakukan dengan cara survey kepada 50 orang pendengar. Berikut adalah perhitungan MOS :

$$(1) \sum_{i=1}^n \frac{x(i) * k}{n}$$

Dimana :

x(i) = Nilai Sample ke i

k = Jumlah bobot

n = Jumlah Pengamatan

3.5 SNR(Signal to Noise Ratio)

Untuk pengukuran nilai energi pada sinyal bicara kita harus melibatkan fungsi window. Hal ini karena dalam pengukuran energi sinyal bicara kita harus menyusunnya dalam frame-frame tertentu. Ini merupakan standar dalam teknologi speech processing, sebab secara umum dalam pengolahan sinyal bicara kita terlibat dengan sinyal dengan durasi yang terlalu panjang bila dihitung dalam total waktu pengukuran. Fenomena ini juga dikenal sebagai *short term speech signal energy*.

Untuk menghitung energi sinyal bicara kita gunakan formulasi dasar seperti berikut:

$$1) E = \sum_{t=0}^T (V(t)w(t))^2$$

dimana: w(m) = merupakan fungsi window seperti *hamming, hanning, bartlett..*

Panjang window dalam hal ini adalah m, untuk durasi dari t=0 sampai t=T akan didapatkan window sebanyak n=T/m apabila tidak ada *overlapping* antara window satu dengan yang lain. Jika terjadi *overlapping* antara window satu dengan yang lain, misalnya sebesar m/2, maka jumlah window dalam satu durasi T adalah sebanyak n = 1 + T/(m/2).

Untuk suatu pengamatan energi pada *frame* ke-k bentuk persamaan (1) menjadi:

$$E_k = \sum_{t=0}^T (V(t)w(k-t))^2 \quad 2)$$

dimana k akan menentukan posisi titik-titik window pada sinyal tersebut, ini juga dikenal sebagai model sliding window.

Dengan menggunakan model *short time measurement* dapat digunakan untuk memilah bagian dari sinyal bicara yang merupakan *voiced* atau *unvoiced*. Sebab pada umumnya *unvoiced speech* memiliki durasi yang lebih pendek. Untuk pengukuran sinyal bicara menggunakan window biasanya dipilih panjang window dengan durasi 10 s/d 20 mili detik. Apabila menggunakan frekuensi sampling sebesar 16 Khz, maka nilainya akan ekuivalen dengan sampel sebanyak 160 sampai 320 sampel setiap frame.

SNR adalah perbandingan (ratio) antara kekuatan sinyal dengan kekuatan derau (*noise level*).

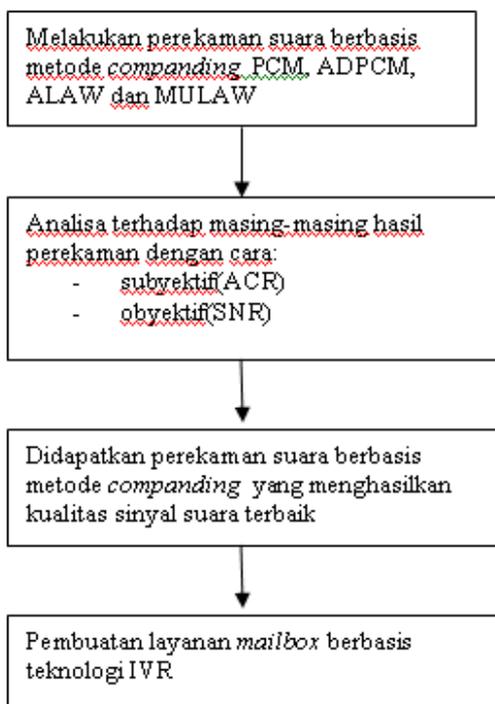
$$SNR = 10 \log_{10} \left(\frac{S(t)^2}{N(t)^2} \right) \quad 3)$$

dimana : S(t) merupakan sinyal sinus tanpa *noise*

N(t) merupakan sinyal *noise*

4.Implementasi dan Hasil Pengujian Sistem

Pada tahap ini, akan dilakukan pengujian kualitas suara dengan menggunakan metode ACR dan SNR, yang bertujuan untuk mengoptimalkan sistem IVR yang dibuat. Alur kerja pembuatan layanan mailbox bisa dilihat pada Gambar 5.



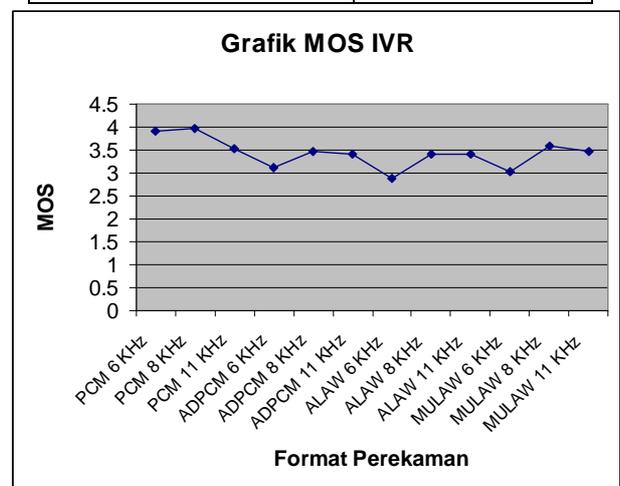
Gambar 5. Alur Kerja Pembuatan Layanan Mailbox

4.1 Hasil Pengujian dengan Metode ACR(Absolute Category Rating)

Pada tahap ini dilakukan proses penentuan tingkat inteligibilitas(kejernihan) suara pada masing-masing format perekaman dengan melakukan survey terhadap 50 responden, dimana masing-masing responden diminta untuk mendengarkan hasil perekaman dengan berbagai format perekaman. Tabel 2 menunjukkan hasil perhitungan nilai rata-rata MOS IVR sedangkan Gambar 6 menunjukkan Grafik MOS IVR

Tabel 2 Hasil Perhitungan Nilai Rata-rata MOS IVR

Format perekaman	Nilai MOS
PCM 6 KHz	3.92
PCM 8 KHz	3.98
PCM 11 KHz	3.52
ADPCM 6 KHz	3.12
ADPCM 8 KHz	3.46
ADPCM 11 KHz	3.4
ALAW 6 KHz	2.88
ALAW 8 KHz	3.4
ALAW 11 KHz	3.42
MULAW 6 KHz	3.04
MULAW 8 KHz	3.6
MULAW 11 KHz	3.46

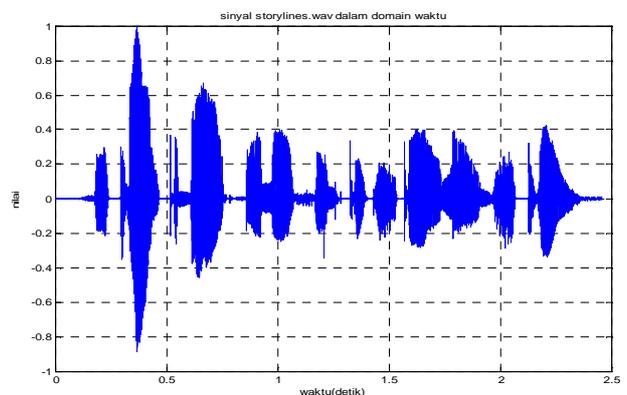


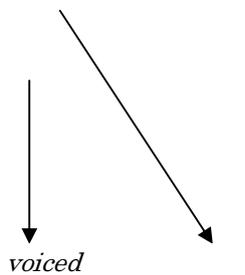
Gambar 6. Grafik MOS IVR

Gambar 6 menunjukkan hasil perhitungan MOS pada masing-masing format perekaman, dimana tingkat inteligibilitas tertinggi didapatkan pada format perekaman PCM dengan *sampling rate* 8 KHz. Pada perbandingan *sampling rate* dari masing-masing format perekaman rata-rata inteligibilitas memiliki nilai tinggi pada *sampling rate* 8 KHz.

4.2 Hasil Pengujian dengan SNR(Signal to Noise Ratio)

Pengujian SNR ini dilakukan dengan cara membandingkan kekuatan sinyal dengan kekuatan *noise*, dimana kekuatan sinyal dilihat dari beberapa sampel bagian *voiced* sedangkan kekuatan *noise* diambil dari beberapa bagian *unvoiced*. Bagian yang menunjukkan *voiced* dan *unvoiced* bisa dilihat pada Gambar 7.





Gambar 7. Bagian *voiced* dan *unvoiced* dari Sinyal Suara

$$SNR = 10 \log_{10} \left(\frac{S(t)^2}{N(t)^2} \right)$$

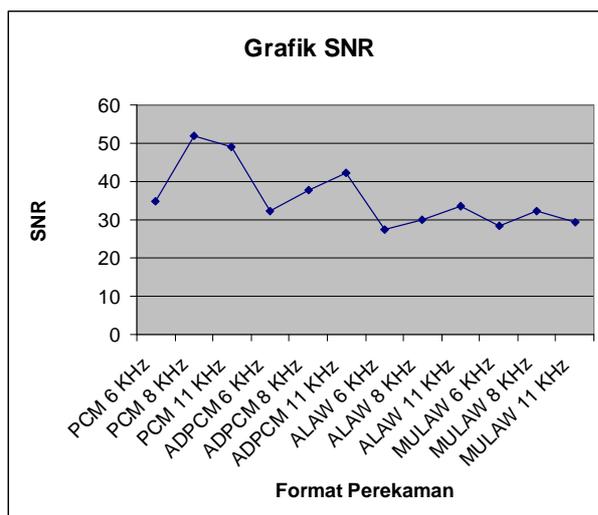
Dimana:

N(t) didapat dari bagian *unvoiced* sedangkan S(t) didapat dari bagian *voiced* yang dikurangi dengan N(t).

Tabel 3 menunjukkan hasil perhitungan SNR sedangkan Gambar 8 menunjukkan Grafik SNR dari berbagai format perekaman.

Tabel 3 Hasil Perhitungan Nilai SNR

Format perekaman	Nilai SNR(dB)
PCM 6 KHz	35
PCM 8 KHz	51.89
PCM 11 KHz	48.9
ADPCM 6 KHz	32.3
ADPCM 8 KHz	37.8
ADPCM 11 KHz	42.3
ALAW 6 KHz	27.3
ALAW 8 KHz	30.1
ALAW 11 KHz	29
MULAW 6 KHz	28.4
MULAW 8 KHz	32.3
MULAW 11 KHz	29.4



Gambar 8. Grafik SNR IVR

Gambar 8 menunjukkan hasil perhitungan SNR pada masing-masing format perekaman, dimana nilai SNR tertinggi didapatkan pada format perekaman PCM dengan *sampling rate* 8 KHz. Pada perbandingan *sampling rate* dari masing-masing format perekaman, rata-rata SNR memiliki nilai tinggi pada *sampling rate* 8 KHz

Dari Gambar 6 dan 8 terlihat bahwa *inteligibility* suara yang diwakili oleh nilai MOS memiliki kaitan erat dengan SNR, dimana jika nilai MOS tinggi maka nilai SNR juga tinggi demikian juga sebaliknya. Hal ini menunjukkan bahwa semakin besar *power* sinyal terhadap *power noise* maka semakin jelas dan jernih suara yang dihasilkan.

Karena dari pengujian MOS dan SNR format perekaman PCM 8 KHz memberikan kualitas sinyal terbaik maka format perekaman ini yang akan digunakan pada layanan *layanan mailbox* yang dibuat.

5. Kesimpulan

1. Tingkat *inteligibilitas* tertinggi didapatkan pada format perekaman PCM dengan *sampling rate* 8 KHz. Sedangkan pada perbandingan *sampling rate* dari masing-masing format perekaman, rata-rata *inteligibilitas* memiliki nilai tinggi pada *sampling rate* 8 KHz.
2. Nilai SNR tertinggi didapatkan pada format perekaman PCM dengan *sampling rate* 8 KHz. Sedangkan pada perbandingan *sampling rate* dari masing-masing format perekaman, rata-rata SNR memiliki nilai tinggi pada *sampling rate* 8 KHz

6. Referensi

- [1] A.Sharma Grover, M.P. Plauche, C.Kuun & E.Barnard, "HIV Health information Access using Spoken Dialogue Systems", ICTD, Qatar, 2009.
- [2] C.Delogu, A.D. Carlo, P.Rotundi & D.Sartori, "Usability Evaluation of IVR Systems with DTMF aand ASR", ICSLP, Australia, 1998
- [3] M.Yuliana, P. Kristalina, "Aplikasi Informasi Nilai Indeks Prestasi Mahasiswa PENS ITS Berbasis Computer Telephony Integration(CTI)", Proceeding of IES, Surabaya, Nopember 2005.

- [4] N.Patel, S.Agarwal, N.Rajput, A.Nanavati,P.Dave&T.Parikh, “*A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India*”, ACM CHI ,USA, 2009.
- [5] P. Kristalina, “*Analisa Tingkat Inteligibilitas Suara pada Layanan Interactive Voice Response dengan Perekaman Berbasis Metode Companding PCM dan ADPCM*”, Proceeding of SNATI, Yogyakarta, Juni 2007.