

## ANALISA TINGKAT INTELIGIBILITAS SUARA PADA LAYANAN INTERACTIVE VOICE RESPONSE DENGAN PEREKAMAN BERBASIS METODE COMPANDING PCM DAN ADPCM

Prima Kristalina

Politeknik Elektronika Negeri Surabaya (PENS)

e-mail: prima@eepis-its.edu

### ABSTRAKSI

Hal yang penting dalam penyiapan layanan Interactive Voice Response adalah pembuatan file database suara yang selalu dibunyikan saat sistim ini diakses. Dua teknik companding yang digunakan pada penelitian ini adalah Linear PCM dan ADPCM. Kedua teknik ini diaplikasikan pada proses perekaman suara dengan dua jenis alat perekam yang berbeda, yaitu carbon microphone dan ribbon microphone. Kualitas hasil perekaman diukur dengan dua cara, yaitu Mean Opinion Score (MOS) untuk mengukur tingkat kejernihan suara dan ukuran file hasil companding. Dari hasil survey terhadap 20 responden didapatkan bahwa tingkat kejernihan hasil perekaman tertinggi adalah menggunakan metode companding A-law dan mu-law dengan frekuensi sampling 11 kHz, dengan nilai MOS 4 dan 3.89 (untuk perekaman dengan ribbon microphone) serta 3.33 dan 3.66 (untuk perekaman dengan carbon microphone). Persentasi kompresi tertinggi adalah dengan metode ADPCM 6 kHz (19,34 %) sedangkan penyusutan terendah dibuat dengan metode A-law dan mu-law 11 kHz (rata-rata 68,16%). Metode A-law dan mu-law mempunyai hasil yang sama karena masih dalam satu kelas Linear PCM, hanya dibedakan pada proses kuantisasinya saja. Hasil perekaman suara semakin jernih jika menggunakan frekuensi sampling tinggi, namun ukuran file hasil companding yang dihasilkan juga semakin besar.

**Kata kunci:** companding, ADPCM, Linear PCM

### 1. PENDAHULUAN

Layanan Interactive Voice Response (IVR) adalah sistim layanan umum yang dapat diakses melalui jalur telepon. Layanan IVR ini banyak diaplikasikan untuk penyampaian informasi dimana user pengakses tidak perlu harus menuju ke lokasi dimana informasi tersebut berada. Sebagai contoh, layanan untuk akses informasi bank, informasi tagihan telepon, informasi rumah sakit dan sebagainya. Meskipun dewasa ini informasi-informasi tersebut sudah dapat diakses via internet, namun perlu diingat jumlah pengguna internet di Indonesia masih lebih sedikit dibandingkan dengan jumlah pengguna telepon, sehingga layanan IVR ini masih dapat diunggulkan.

Untuk membuat server IVR, perlu disediakan sebuah PC yang dilengkapi dengan Voice Processing Card, yang terhubung dengan jalur telepon.

Hal yang penting dalam penyiapan sebuah layanan IVR adalah pembuatan file data base suara yang akan selalu dibunyikan saat diakses. Pembuatan database suara ini dilakukan melalui proses perekaman suara. Proses perekaman suara dapat dilakukan dengan dua cara : melalui fasilitas software perekaman suara dan fasilitas software yang disediakan oleh Voice Processing Card.

Yang perlu diperhatikan dalam perekaman suara untuk layanan IVR ini adalah bahwa format perekaman harus mengikuti standart suara yang digunakan dalam sistim teleponi (frekuensi sampling sekitar 6 kHz s/d 11 kHz). Banyak software pemrosesan sinyal yang menyediakan fasilitas

perekaman, namun umumnya digunakan untuk fasilitas sound audio sehingga tidak semua dapat digunakan dalam sistim teleponi. Dalam penelitian ini digunakan prinsip perekaman dengan format companding ADPCM, a-law dan mu-law. Ketiga macam format ini dapat dikenali dan digunakan untuk layanan IVR.

### 2. INTELIGIBILITAS SUARA

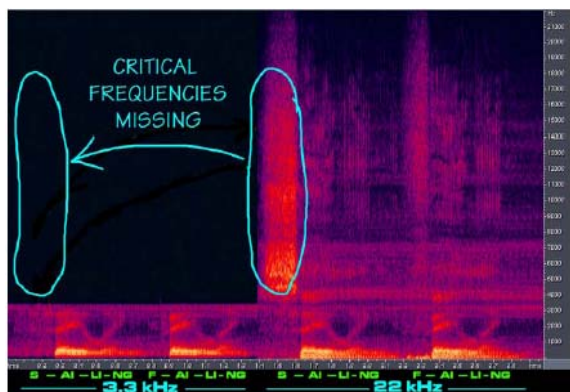
Inteligibilitas suara adalah ukuran untuk menyatakan tingkat kephahaman maksud dari suara yang diucapkan oleh pembicara.<sup>1</sup> Biasanya dinyatakan dalam persen informasi yang dapat dipahami. Ada beberapa faktor yang mempengaruhi tingkat inteligibitas suara, diantaranya adalah audibilitas, kejernihan (*clarity*), kecepatan bicara, aksan dan artikulasi pembicara. Audibilitas dalam hal ini termasuk gaung (*reverberation*) dan echo. *Clarity* dapat berkurang karena faktor-faktor: distorsi amplitudo yang diakibatkan oleh hardware atau peralatan elektronik, distorsi frekuensi yang diakibatkan oleh hardware/elektronik maupun lingkungan akustik serta distorsi berdasarkan domain waktu yang menyebabkan pantulan (refleksi) dan gaung/gema.

Crandall pada 1917 menjelaskan bahwa untuk mengidentifikasi suatu kata, tanpa mempedulikan jenis vokalnya, faktor yang berpengaruh adalah pada konsonannya, terutama dalam artikulasinya.<sup>2</sup>

Untuk menjelaskan maksud di atas, diberikan kalimat sebagai berikut, "bawakan saya kasur", tentu beda maknanya dengan "bawakan saya kapur". Cara

orang melafalkan sesuatu juga bisa membuat kesalahan dalam pendengaran. Contoh, orang Jawa mengatakan “lara (diucapkan loro, seperti o dalam doa)”, yang berarti sakit dan “loro (diucapkan loro, seperti o dalam soto atau bakso), yang berarti dua. Ke-dua kata ini tentu akan mempunyai arti yang sama sekali berlainan jika salah pengucapannya.

Kata-kata dalam bahasa Indonesia lebih banyak memuat unsur vokal dari pada konsonan, berbeda dengan kata-kata dalam bahasa Inggris, Perancis maupun Jerman. Sebenarnya unsur vokal tidak terlalu banyak berpengaruh dalam bandwidth suara, karena masih berada di bawah 3 kHz, sedangkan unsur konsonan bisa di atas 3 kHz. Unsur konsonan ini akan sangat kelihatan pengaruhnya saat suara melewati proses sampling. Gambar 1 memperlihatkan bentuk suara untuk kata “sailing” dan “failing”.



Gambar 1. Bentuk suara “sailing” dan “failing”

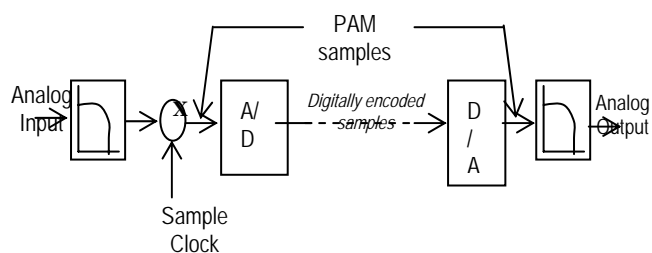
Pengukuran inteligibilitas suara dapat dilakukan dengan metode *Mean Opinion Score* (MOS).<sup>3</sup> Metode ini merupakan standart pengukuran kualitas suara terutama pada jalur telepon. MOS biasanya diukur pada rangkaian penerima. Metode pengukuran MOS adalah dengan cara user mendengarkan kualitas suara yang diterima, selanjutnya memberikan opini berupa nilai/score tertentu tergantung dari jernih tidaknya kualitas suara yang diterima. Selanjutnya score yang didapat dari para user tersebut akan dirata-rata secara matematis untuk menentukan kualitas suara tersebut. Ada 5 macam score yang dapat diberikan oleh user , yaitu: (1) *bad*; (2) *poor*; (3) *fair*; (4) *good*; (5) *excellent*. Skema pengukuran model MOS ini dikenal sebagai metode pengukuran subyektif, karena bergantung dari opini masing-masing subyek.

### 3. PROSES COMPANDING DENGAN METODE PCM

Untuk dapat didengar oleh user penerima pada jarak yang jauh, suara manusia yang dihasilkan pada pengirim harus melalui proses modulasi. Proses modulasi yang digunakan pada teknik teleponi yang umum adalah PCM (*Pulse Code Modulation*), dimana pada proses ini pada sisi pengirim terdapat

tiga sub proses yaitu: sampling, kuantisasi, dan coding. Proses sampling merupakan suatu urutan pengambilan sample-sampel dari sinyal suara yang akan ditransmisikan. Berdasarkan teorema Nyquist<sup>4</sup> dijelaskan bahwa proses sampling suara manusia yang ideal adalah jika frekuensi sampling-nya kurang lebih sama dengan dua kali frekuensi suara manusianya ( $f_s = 2f_i$ ). Jika frekuensi sampling kurang atau lebih dari dua kali frekuensi suara maka akan terjadi *aliasing*, dimana ada sebagian informasi yang hilang atau berlebih.

Ada 2 jenis proses PCM yang digunakan dalam perekaman suara pada penelitian ini: *Linear PCM/Uniform PCM* dan *Adaptive Differential PCM* (ADPCM). Prinsip *Linear PCM* ditunjukkan pada gambar 2.



Gambar 2. Prinsip Linear PCM

Pada *Linear PCM* seluruh sinyal input analog (dalam hal ini suara manusia) diambil per-sampel dengan frekuensi sampling tertentu (sekitar 8 kHz ), selanjutnya akan dikuantisasi dan dikodekan dengan metode companding mu-law. Companding mu-Law bertujuan untuk mengkompresi data informasi sekaligus mendapatkan kode 8 bit-nya. Definisi mu-law adalah:

$$F_{\mu}(x) = \text{sgn}(x) \frac{\ln(1 + \mu |x|)}{\ln(1 + \mu)}$$

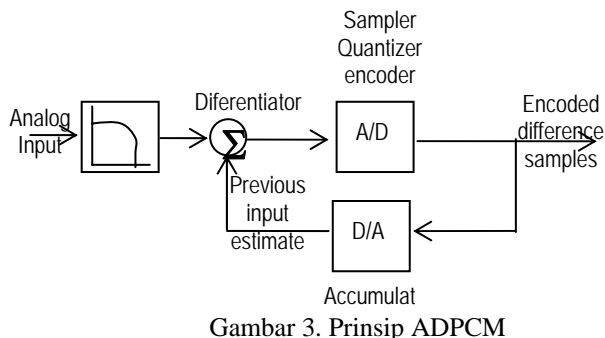
dimana:

$x$  = amplitud sinyal input ( $-1 \leq x \leq 1$ )

$\text{sgn}(x)$  = polaritas  $x$

= parameter yang digunakan untuk menyatakan nilai kompresi.

Pada *Adaptive Differential PCM*, sample-sampel yang berbeda terlebih dahulu disimpan sementara pada rangkaian *sample and hold*. Selanjutnya melalui rangkaian *analog subtractor* dibandingkan antara sample yang tiba terlebih dahulu dengan sample berikutnya. Perbedaan antar sample tadi yang akan dikuantisasi dan dikodekan untuk proses transmisi. Perbedaan dengan Linear PCM adalah bahwa pada *Adaptive Differential PCM* ini ada proses *feedback* untuk mendapatkan perbedaan nilai sample. Proses *Adaptive Differential PCM* (ADPCM) ditunjukkan pada gambar 3.



Gambar 3. Prinsip ADPCM

#### 4. PERENCANAAN SISTIM

Proses perekaman suara dilakukan pada dua jenis tools yang berbeda. Tool pertama adalah menggunakan *carbon microphone*, yang terpasang pada pesawat telepon. Dengan pemrograman menggunakan C++ untuk IVR, dilakukan perekaman dengan 6 jenis teknik *companding* yang berbeda, yaitu: ADPCM 6 kHz, ADPCM 8 kHz, A-law 8 kHz, A-law 11 kHz, mu-law 8 kHz dan mu-law 11 kHz. Perlu diperhatikan, ke enam teknik *companding* ini berada pada jangkauan bandwidth 6 kHz – 11 kHz, yang merupakan batasan bandwidth yang dikenali oleh Voice Processing Board yang terpasang pada server IVR. Tool kedua adalah perekaman menggunakan *ribbon microphone* yang terpasang pada PC (microphone ini biasa digunakan untuk aplikasi *signal processing* maupun VoIP). Dengan mengaplikasikan perangkat lunak Vox Studio Rel.3.0.94, perekaman dilakukan pada teknik *companding* yang sama, yaitu ADPCM 6 kHz, ADPCM 8 kHz, A-law 8 kHz, A-law 11 kHz, mu-law 8 kHz dan mu-law 11 kHz. Hasil perekaman disimpan dalam bentuk file .WAV. Masing-masing file tersebut diakses oleh user penerima saat dia terhubung dengan server IVR. User selanjutnya memberikan score terhadap kejernihan masing-masing suara yang diperdengarkan.

Score yang didapat diukur rata-ratanya (*mean*), dimana persamaan untuk nilai rata-rata diberikan sebagai:

$$mean = \frac{\sum_{i=1}^m x(i).i}{n}$$

Dimana:

x(i) = nilai sampel ke-i

i = jumlah bobot

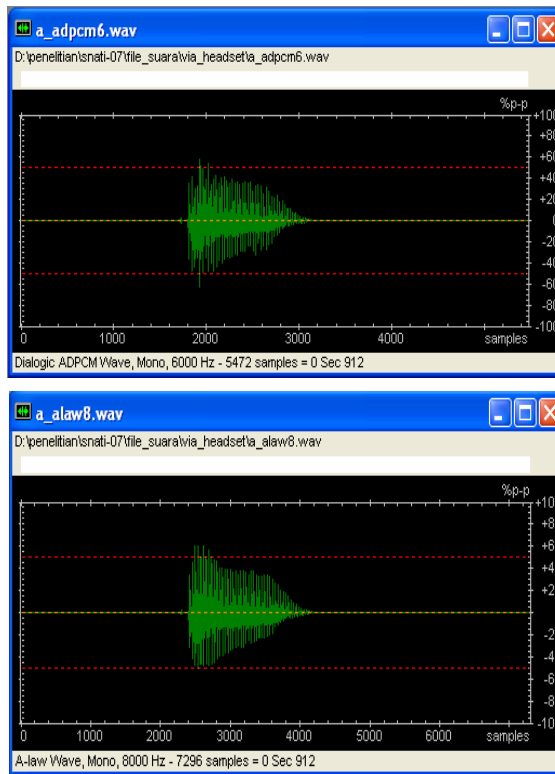
n = jumlah sampel

Selain dapat diukur berdasarkan metode *Mean Opinion Score* (MOS), akan diukur pula besarnya kompresi untuk masing-masing file yang telah mengalami proses *companding*. Besarnya file hasil kompresi ini menunjukkan besarnya prosentase pemampatan file pada setiap proses *companding*.

#### 5. HASIL PENGUKURAN DAN ANALISA

Dari pengukuran yang telah dilakukan terhadap file database suara dengan berbagai metode *companding*, dapat ditentukan tiga jenis analisa: 1) Bentuk sinyal hasil *companding*, 2) *Mean Opinion Score*, 3) Ukuran File hasil *companding*.

##### 5.1 Bentuk Sinyal Hasil Companding

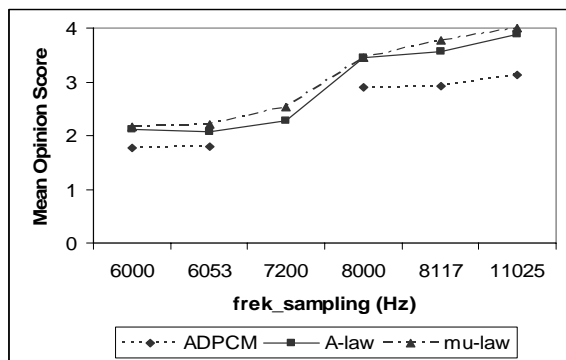


Gambar 4. Bentuk sinyal hasil companding ADPCM dan A-law

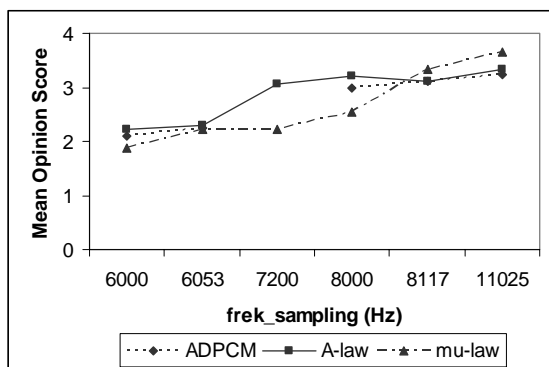
Kedua gambar di atas mewakili 12 gambar yang lain, merupakan bentuk perekaman suara “a” dengan frekuensi input sampling 8 kHz, 16 bit linier PCM, diformat dengan metode *companding* ADPCM 6 kHz dan A-law 8 kHz. Level sinyal suara lebih tinggi terdapat pada metode A-law dibandingkan dengan ADPCM.

##### 5.2 Mean Opinion Score (MOS)

Pengamatan dilakukan dengan men-survey 20 responden untuk mendengarkan dua jenis file, yaitu perekaman melalui headset PC (dengan *ribbon microphone*) dan melalui pesawat telepon (dengan *carbon microphone*). Hasil rata-rata nilai MOS ditunjukkan pada gambar 5.



a) Perekaman dengan ribbon microphone



b) Perekaman dengan carbon microphone

Gambar 5. Mean Opinion Score untuk file a.wav

Dari hasil pengukuran didapatkan bahwa perekaman dengan metode companding ADPCM mendapatkan nilai terburuk dibandingkan dengan metode A-law dan mu-law. Semakin tinggi frekuensi sampling yang digunakan untuk masing-masing metode, semakin bagus score yang diberikan. Perekaman dengan carbon microphone mempunyai nilai MOS lebih rendah dibandingkan dengan ribbon microphone. Perlu diingat bahwa metode *Mean Opinion Score* adalah metode yang subyektif, sangat bergantung kepada kondisi penfengaran masing-masing responden serta peralatan *speaker* yang disediakan.

### 5.3 Ukuran file hasil Companding.

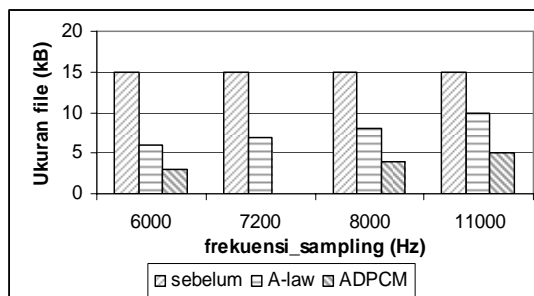
Setelah melalui proses companding, ukuran file asal menyusut sesuai dengan metode companding yang digunakan. Pada model pengamatan ini diukur jumlah persen penyusutan file serta ukuran file sesungguhnya setelah mengalami proses companding. Tabel 1 menunjukkan besaran 4 macam file (15 kB, 58 kB, 67 kB dan 79 kB) setelah mengalami proses companding dengan metode ADPCM, A-law dan mu-law.

Dari table dapat dianalisa bahwa penyusutan tertinggi adalah jika perekaman dilakukan dengan metode ADPCM 6 kHz (yaitu rata-rata 19,34 %) sedangkan penyusutan terendah dibuat dengan metode A-law dan mu-law 11 kHz (rata-rata 68,16%).

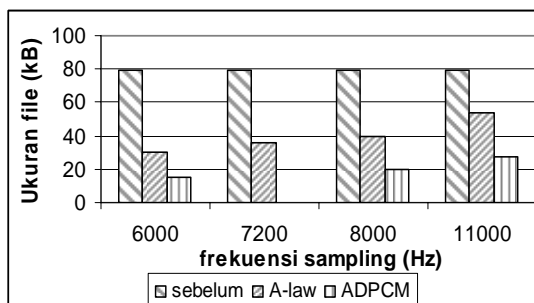
Tabel 1. Nilai penyusutan ukuran file berdasarkan metode Companding ADPCM, A-law dan mu-law

Ukuran file Metode Companding	15 kB	58 kB	67 kB	79 kB	Rata-rata % Comp (%)
	% Comp (%)	% Comp (%)	% Comp (%)	% Comp (%)	
adpcm_6000	20.00	18.96	19.40	18.98	19.34
adpcm_8000	26.00	25.86	25.37	25.32	25.64
adpcm_11000	33.00	34.48	34.33	34.18	34.00
a_low_6000	40.00	37.93	37.31	37.97	38.30
a_low_7200	46.67	44.83	44.78	45.57	45.46
a_low_8000	53.33	50.00	50.75	50.63	51.18
a_low_11000	66.67	68.97	68.66	68.35	68.16
mu_low_6000	40.00	37.93	37.31	37.97	38.30
mu_low_7200	46.67	44.83	44.78	45.57	45.46
mu_low_8000	53.33	50.00	50.75	50.63	51.18
mu_low_11000	66.67	68.97	68.66	68.35	68.16

Penyusutan ukuran file untuk ukuran awal 15 kB dan 79 kB ditunjukkan pada gambar 7, dimana dengan metode companding ADPCM didapatkan ukuran penyusutan terbesar.



a. Ukuran file asal 15 kB



b. Ukuran file asal 79 kB

Gambar 7. Perbandingan ukuran file sebelum dan sesudah proses companding

## 6. KESIMPULAN

Salah satu masalah yang perlu diperhatikan di dalam membuat sebuah sistim layanan informasi berbasis *Interactive Voice Response* adalah model perekaman suara yang nantinya akan dibunyikan saat sistim tersebut diakses oleh user. Penelitian ini telah melakukan perekaman dengan 3 jenis metoda companding yang berbeda dan 2 jenis media perekaman, yaitu *ribbon microphone* dan *carbon microphone* dan didapatkan kesimpulan sebagai berikut:

1. Level sinyal suara yang diproses dengan metode companding ADPCM lebih rendah dibandingkan dengan level sinyal suara hasil companding A-law dan mu-law.

2. Score tertinggi diberikan pada hasil perekaman dengan metode companding A-law (4) dan mu-law 11 kHz (3,89), terendah pada hasil perekaman dengan ADPCM 6 kHz.
3. Dengan metode companding yang sama, perekaman dengan media ribbon microphone mempunyai score yang lebih tinggi dibandingkan dengan carbon microphone.
4. Penyusutan ukuran file terbanyak diperoleh pada perekaman dengan metode ADPCM 6 kHz (yaitu rata-rata 19,34 %) sedangkan penyusutan terendah dibuat dengan metode A-law dan mu-law 11 kHz (rata-rata 68,16%).
5. Metode A-law dan mu-law mempunyai hasil yang sama, karena kedua jenis metode ini masih satu grup, yaitu grup Linear PCM.
6. Kualitas kejernihan suara berbanding lurus dengan besar amplitude sinyal hasil perekaman, besar frekuensi sampling dan besar ukuran file yang dihasilkan setelah proses perekaman.

#### **PUSTAKA**

- [1] Larry Young, *Measuring Voice Quality in Hybrid Network*, Technology Information, Telecommunication, November 2000
- [2] Mark A. Miller, *Do you hear what I hear? : Measuring Toll Quality*, July 2005
- [3] International Telecommunications Union—Telecommunications Standardization Sector (ITU-T), *Methods of Subjective Determination of Transmission Quality*, designated P.800
- [4] John Bellamy, *Digital Telephony*, John Wiley & Sons, 2005
- [5] Endah Kurniawati, *Aplikasi Studi Islam Anak-anak dengan Menggunakan Card Dialogic 4*, Tugas Akhir, PENS, 2005.