

Penyandi Persepsi Isyarat Audio Berdasar pada Model Modulasi Frekuensi (FM)

Bondhan Winduratna¹, Adhi Susanto², Risanuri Hidayat³

Abstract—Data compression or source coding has been playing an important role in many areas in daily life. There are several audio encoders available with various methods. Each audio encoder works at a certain range bitrate depending on the signal model used and the method implemented. Based on literature reviews it is to know that there are only a view of audio encoder working at bit rate below 16 kbit/s, especially for music signals.

The togetherness in having harmonically components between the audio signals and frequency modulated signals is the basic idea of proposing the perceptual encoder based on FM signal model. In estimating and extracting FM parameters, the psychoacoustic model is used to exploit the perceptual nature of humans ears. The errors of FM modeling is distributed below the masking curve of psychoacoustic model. The developed encoder allows sending music signals on the GSM channel.

The developed encoder has four cascaded main processes, namely sinusoid components extraction, harmonic components separation, FM parameters estimation, and FM parameter encoding. The simulation results show the proposed method is successful to encode some audio signal types.

Intisari—Kompresi data atau penyandian sumber telah dan masih memainkan sebuah peran penting dalam berbagai area di kehidupan sehari-hari. Setelah mengalami puluhan tahun pengembangan terdapat beberapa penyandi audio dengan metode yang berbeda. Setiap penyandi audio bekerja pada cakupan pesat bit tertentu, tergantung pada jenis model isyarat yang digunakan dan metode yang dipakai. Terdapat hanya sedikit penyandi audio yang bekerja pada pesat bit di bawah 16 kbit/s, terutama untuk isyarat musik.

Kebersamaan dalam mempunyai komponen harmonik antara isyarat musik dan isyarat yang termodulasi frekuensi, merupakan ide dasar pembuatan penyandi persepsi berdasar pada model isyarat FM. Dalam mengestimasi dan mengekstrasi parameter FM, sebuah model psikoakustik digunakan untuk mengeksploitasi sifat persepsi telinga manusia. Penyandi yang diusulkan memungkinkan untuk mengirim isyarat musik pada saluran GSM.

Penyandi yang diusulkan mempunyai empat proses pokok, yaitu ekstraksi komponen sinusoid, separasi komponen harmonik, estimasi parameter FM, dan penyandian parameter FM. Hasil-hasil simulasi memperlihatkan bahwa penyandi yang diusulkan telah sukses berhasil menyandikan beragam jenis isyarat audio.

Kata kunci :isyarat, harmonik pemodelan, modulasi frekuensi

I. PENDAHULUAN

Untuk berbagai macam aplikasi pada bidang telekomunikasi, hiburan elektronis, dan sebagainya, telah

^{1,2,3}Dosen, Departemen Teknik Elektro dan Teknologi Informasi, Fakultas Teknik UGM, Jl. Grafika 2, Yogyakarta 55281 INDONESIA (tlp: 0274 552305; e-mail: windurat@ugm.ac.id)

tersedia berbagai penyandi audio yang berbeda satu dengan lain dalam hal kualitas, pesat bit, kompleksitas, dan waktu latensi. Rangkuman studi literatur berbagai penyandi tersebut adalah sebagai berikut.

Penyandi yang bekerja dalam ranah waktu biasanya berdasar algoritme prediksi linear. Korelasi antar sampel isyarat audio direduksi dengan penerapan prediksi linear [1].

Adaptive Differential Pulse-Code Modulation (ADPCM) [2] mempunyai koefisien prediktor yang adaptif terhadap dinamika isyarat. Pesat bit yang ditawarkan berkisar dari 16 kbit/s sampai 64 kbit/s dengan kualitas mendekati kualitas PCM 64 kbit/s.

Code Exited Linear Prediction (CELP) [3] merupakan pengembangan dari ADPCM dengan penambahan *adaptive postfilter* untuk mereduksi *irrelevance*. Pesat bit yang dicapai pada 16 kbit/s dengan kualitas mendekati ADPCM 32 kbit/s.

Full-Rate Speech Transcoding [4] merupakan penyandi yang dipakai pada jaringan GSM. Penyandi dengan pesat bit 13 kbit/s ini berbasis pada algoritme CELP dengan pengembangan pada untai prediksinya, prediksi waktu singkat dan prediksi waktu panjang dipakai untuk menambah kinerja penyandi.

Ketiga penyandi dengan prediksi linear di atas secara spesifik dikembangkan untuk isyarat tutur dan penggunaannya diprioritaskan pada sistem komunikasi telepon. Untuk isyarat musik, selanjutnya dikembangkan penyandi berbasis transformasi. Korelasi antar sampel isyarat dihilangkan dengan merepresentasikan isyarat melalui koefisien dari transformasi waktu ke frekuensi.

Penyandi Audio 7 kHz [5] merupakan penyandi yang dirancang untuk isyarat tutur dan musik dengan pesat bit dari 56 kbit/s sampai 148 kbit/s. Algoritme yang dipakai adalah *subband coding*. Penyandi ini dikembangkan untuk konferensi Audio-Visual pada jaringan ISDN. Kualitas yang dicapai cukup baik untuk isyarat tutur dan musik dengan lebar bidang 50 Hz – 7 kHz.

Penyandi audio MPEG1 [6], MPEG2 [7],[8] dibuat untuk mencukupi kebutuhan penyandi audio dengan kualitas CD. Seperti halnya pada penyandi G722, penyandi ini juga berbasis pada *transforms coding* atau *subband coding*, hanya saja penyandi ini menggunakan model psikoakustik untuk mengendalikan resolusi kuantisasi koefisien *filter bank*, sehingga mempunyai kinerja yang lebih baik. Pesat bit yang ditawarkan berkisar dari 128 kbit/s sampai 384 kbit/s.

Penyandi audio MPEG4 [9], [10] dibuat dengan menggunakan gabungan teknik-teknik penyandian audio yang optimal di masing-masing daerah kerja yang meliputi aspek jenis isyarat audio dan aspek besarnya pesat bit. Secara garis besar algoritme yang dipakai adalah HVXC dan CELP untuk isyarat tutur, MPEG-2 AAC untuk isyarat musik, dan

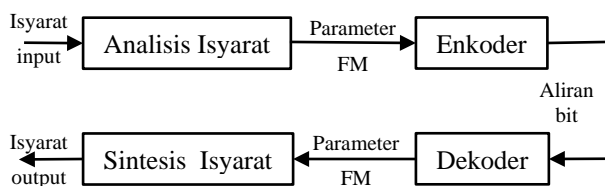
algoritme penyandi parametrik untuk gabungan isyarat tutur dan musik. Dengan demikian, penyandian ini juga dapat bekerja dari pesat bit 6 kbit/s sampai 384 kbit/s dan cocok untuk aplikasi multimedia dengan cakupan yang luas dengan kualitas dari setara GSM sampai CD.

Dari studi literatur diketahui bahwa terdapat daerah pesat bit di bawah 32 kbit/s yang secara eksplisit belum banyak dieksploitasi terutama untuk penyandian isyarat musik. Pada makalah ini dicoba dibuat penyandi persepsi untuk isyarat audio yang bekerja pada pesat bit di bawah 16 kbit/s.

Penyandi yang akan diusulkan akan berkerja berdasar konsep parameterik berbasis model FM [11]. Prinsip dasar penyandi parameterik adalah menganalisis isyarat *input* untuk mendapatkan parameter-parameter model isyarat dan selanjutnya menyandikan parameter-parameter model menjadi aliran bit dan dikirim ke penerima. Pada sisi penerima aliran bit didekodekan menjadi parameter model dan kemudian direkonstruksi kembali ke isyarat audio asli melalui proses sintesis isyarat.

Terdapat dua fakta pengamatan pada isyarat musik yang mengilhami pembuatan penyandi. Pertama adanya kesamaan sifat harmonis yang dimiliki oleh mayoritas isyarat musik dan isyarat FM. Kedua, isyarat musik mempunyai spektrum yang stasioner dalam potongan waktu sesaat. Sifat harmonis dari isyarat musik diekspresikan melalui pemodelan isyarat FM. Fakta kedua dituangkan dalam cara kerja penyandi yang prediktif di ranah frekuensi. Selain itu, untuk meningkatkan kualitas dan efisiensi pemodelan isyarat FM, model psikoakustik diandalkan untuk menentukan konvergensi kesempurnaan pemodelan. Penyandi yang dikembangkan mempunyai empat tahapan proses pokok, yaitu ekstraksi komponen sinusoid, separasi komponen harmonik, estimasi parameter FM, dan penyandian parameter FM.

Penyandi persepsi berdasar modulasi frekuensi ini diharapkan dapat menjadi alternatif penyandi isyarat audio pada pesat bit di bawah 16 kbit/s. Secara garis besar, sistem penyandian dapat digambarkan seperti dalam Gbr. 1.



Gbr. 1 Konsep penyandi parametrik berbasis Modulasi Frekuensi (FM).

Untuk mendapatkan parameter FM dari isyarat audio perlu dilakukan analisis isyarat yang terbagi dalam tiga tahapan. Tahapan pertama adalah ekstraksi komponen-komponen sinusoid. Tahapan kedua adalah separasi komponen harmonik dengan komponen in-harmonik. Tahapan ketiga merupakan proses parameterisasi komponen harmonik menjadi parameter FM.

A. Model Isyarat Harmonik

Isyarat audio merupakan salah satu media pembawa informasi dalam komunikasi suara interaktif antar peserta

komunikasi. Isyarat audio berupa fungsi dengan perubah bebas t dan secara umum dapat dirumuskan (1), yang merupakan jumlah sinusoid kompleks ditambah derau Gaussian putih $z(t)$.

$$x(t) = \sum_{k=1}^N A_k e^{j(\omega_k t + \theta_k)} + z(t) \quad (1)$$

Parameter ω_k , A_k dan θ_k adalah frekuensi, amplitude dan fase. Untuk isyarat yang riil tanpa derau, isyarat audio $x(t)$ dapat dimodelkan dengan isyarat gabungan dari beberapa sinusoid riil tanpa disertai derau $z(t)$.

$$x(t) = \sum_{k=1}^N A_k \cos(\omega_k t + \theta_k) \quad (2)$$

Berbeda dengan isyarat audio, mayoritas isyarat musik mempunyai pola sebaran frekuensi yang harmonis [12][13], sehingga di ranah waktu isyarat musik berlalu secara periodis. Menurut deret Fourier, sebuah isyarat periodis $x(t)$ dapat diaproksimasi dengan isyarat $\hat{x}(t)$ yang tersusun dari sinusoid-sinusoid dengan frekuensi kelipatan bulat frekuensi dasar ω_0 . Isyarat $\hat{x}(t)$ sering disebut isyarat harmonis dan secara matematis dapat diformulasikan dengan (3).

$$\hat{x}(t) = \sum_{n=0}^N A_n \cos(n\omega_0 t + \theta_n), \quad (3)$$

dengan A_n adalah amplitude sinusoid ke- n , ω_0 adalah frekuensi dasar dan θ_n adalah fase sinusoid ke- n . Untuk $n = 0$ disebut komponen DC dan untuk $n \geq 1$ disebut harmonik ke- n . Khusus untuk $n = 1$ disebut komponen dasar karena mempunyai frekuensi dasar ω_0 dan sinusoid untuk $n > 1$ disebut (komponen) harmonik atau *partial tones*.

B. Estimasi Frekuensi

Estimasi frekuensi merupakan salah satu langkah yang penting dalam penyandi berbasis model modulasi frekuensi. Sampai saat ini sudah tersedia berbagai estimator frekuensi yang mungkin dapat diadopsi untuk keperluan penelitian. Setelah melalui seleksi dengan pertimbangan aspek akurasi dan kecepatan komputasi dipilih estimator sinusoid tunggal dari Kay [14].

Selain parameter frekuensi, sebuah sinusoid mempunyai parameter yang lain, yaitu parameter amplitude dan fase. Kedua parameter ini baru dapat diestimasi setelah parameter frekuensi selesai diestimasi. Untuk sinusoid riil $x(n) = A_0 \cos(\omega_0 n + \theta_0)$, parameter amplitude dan fase dapat diestimasi dengan mengkorelasikan silang antara potongan isyarat riil $x(n)$ dengan sinusoid acuan yang mempunyai frekuensi terestimasi $\hat{\omega}_0$ dan parameter fase yang diketahui.

$$\hat{x}(n) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cos(\hat{\omega}_0 n) \quad (4)$$

dengan \hat{A}_0 merupakan hasil estimasi amplitude dan $\hat{\theta}_0$ merupakan hasil estimasi fase. Dengan demikian, untuk mengesktraksi sebuah sinusoid diperlukan estimasi frekuensi dari Kay dan diikuti hitungan korelasi silang. Untuk

mengeskraksi beberapa sinusoid, algoritme ekstraksi sinusoid tunggal di atas dipekerjakan secara iteratif [15].

C. Ekstraksi Komponen Harmonik

Besaran frekuensi fundamental dan komponen harmonik merupakan dua besaran yang tidak dapat saling terpisahkan. Estimasi frekuensi fundamental kadang memerlukan frekuensi-frekuensi harmonik dan penentuan frekuensi harmonik juga memerlukan frekuensi fundamental.

Masalah dasar estimasi frekuensi fundamental adalah frekuensi dasar yang berubah terhadap waktu dari isyarat yang tersedia. Pendekatan atau penyederhanaan yang dilakukan untuk mendasari pengembangan metode estimasi frekuensi fundamental adalah:

- isyarat dianggap kuasi harmonik sehingga periode isyarat atau frekuensi fundamental dapat dianggap eksis,
- frekuensi fundamental menempati daerah cakupan tertentu sehingga dimungkinkannya adanya toleransi kesalahan estimasi,
- laju perubahan frekuensi fundamental dianggap cukup lambat,

Dari banyak metode estimasi frekuensi fundamental yang tersedia metode pengelompokan sinusoid-sinusoid yang harmonis. Metode ini memperkirakan letak frekuensi fundamental berdasar pengelompokan komponen harmonik dari komponen-komponen sinusoid yang telah berhasil diekstraksi. Selain itu metode ini dapat mempertimbangkan adanya kesalahan pada estimasi frekuensi sinusoid seperti pada yang dipakai dalam [16].

D. Teknik Sintesis Frekuensi Modulasi

Penggunaan Teknik modulasi frekuensi pertama kali diusulkan untuk pengiriman isyarat audio lewat gelombang radio pada tahun 1936 [17]. Prinsip kerja modulasi frekuensi adalah memvariasi frekuensi sinusoid isyarat pembawa $A_c \sin(2\pi f_c t)$ dengan isyarat pemodulasi, misalnya sinusoid $A_m \sin(2\pi f_m t)$. Hasil modulasi frekuensi ini merupakan adalah isyarat $x(t)$ pada (5),

$$x(t) = A_c \sin[2\pi f_c t + I \sin(2\pi f_m t)] \quad (5)$$

dengan I merupakan indeks modulasi,

$$I = \frac{2\pi K_t A_m}{\omega_m} = \frac{A_m K_t}{f_m}$$

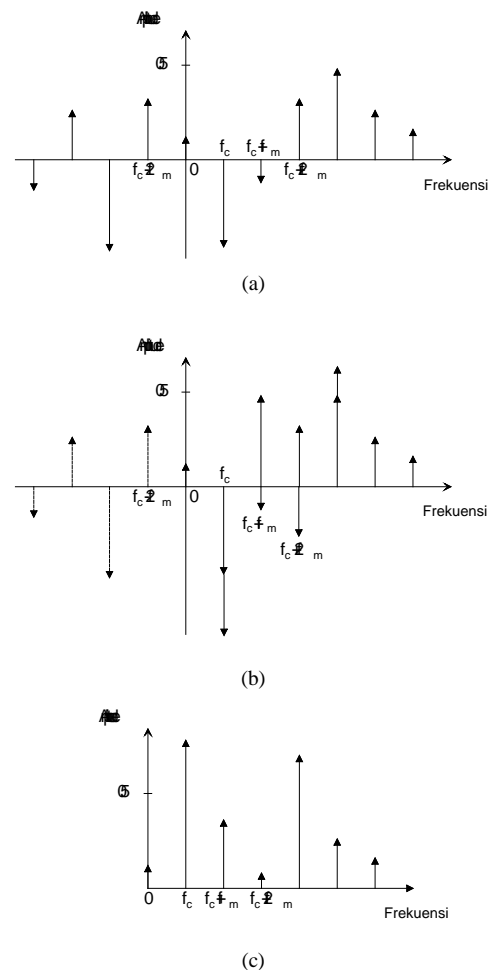
dan K_t adalah konstanta deviasi frekuensi. Persamaan gelombang sinusoid dalam sinusoid pada isyarat $x(t)$ merupakan sebuah persamaan diferensial yang dapat diselesaikan dengan ekspansi deret. Melalui pertolongan fungsi Bessel jenis pertama, isyarat $x(t)$ pada (5) dapat ditulis sebagai deret Fourier.

$$x(t) = \sum_{k=-\infty}^{\infty} J_k(I) A_c \sin[(2\pi f_c + k2\pi f_m)t] \quad (6)$$

dengan $J_k(I)$ merupakan fungsi Bessel jenis pertama dengan variabel bebas indeks modulasi I .

$$J_k(I) = \sum_{r=0}^{+\infty} \frac{(-1)^r \left(\frac{I}{2}\right)^{k+2r}}{r! I(k+r+1)}$$

Di ranah frekuensi isyarat $x(t)$ berupa spektrum garis yang harmonis yang lebar bidangnya dapat mencapai tak berhingga tergantung intensitas indeks modulasi. Lebarnya bidang spektrum isyarat FM yang lebar justru menawarkan alternatif yang sangat menarik untuk sintesis isyarat musik. Spektrum isyarat musik yang kompleks dapat diproduksi hanya dengan sejumlah parameter yang relatif sedikit. Untuk meraih derajat kompleksitas spektrum, Chowing [18] menempatkan frekuensi pembawa f_c sama atau lebih rendah dibanding frekuensi modulasi f_m . Hal ini berkebalikan dengan penggunaan teknik modulasi frekuensi pada telekomunikasi. Bagian spektrum yang terletak di sisi negatif akan tercermin kembali sisi positif. Gbr. 2(a) merupakan ilustrasi pembangkitan spektrum isyarat FM menurut (2) dengan frekuensi pembawa f_c yang sama tinggi dengan frekuensi modulasi f_m . Pada Gbr. 2(b) dapat dilihat bahwa garis spektrum frekuensi negatif akan direfleksikan balik dengan polaritas magnitude yang terbalik ke arah frekuensi positif.



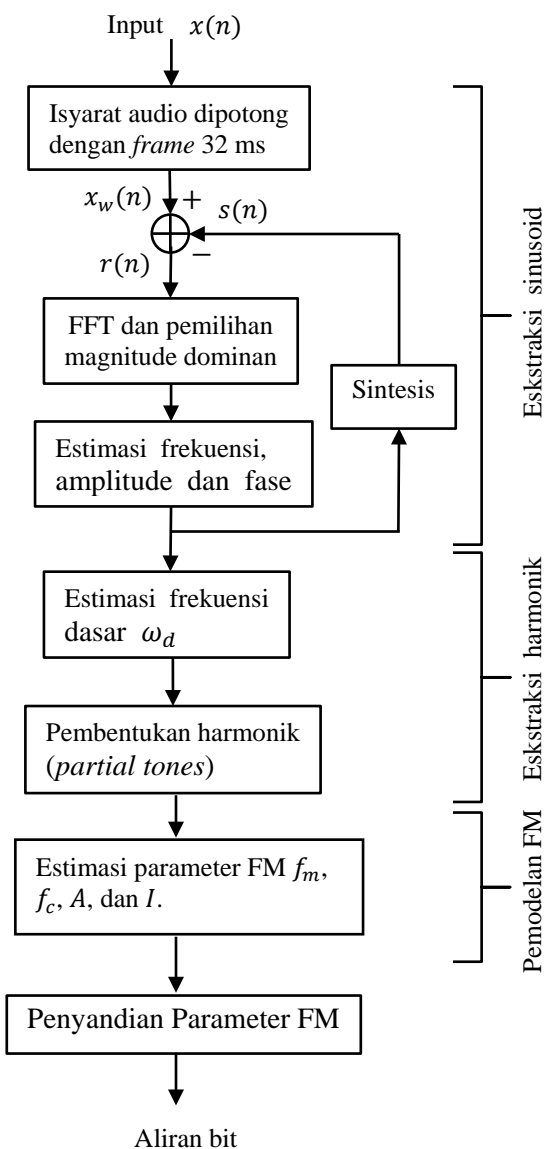
Gbr. 2 (a) Spektrum garis modulasi frekuensi dengan garis spektrum negatif, (b) Garis spektrum negatif direfleksikan ke arah frekuensi positif, (c) Spektrum garis modulasi frekuensi.

Selain mengalami proses pencerminan, komponen spektrum frekuensi negatif juga akan dijumlah dengan komponen spektrum positif pada frekuensi yang sama. Gbr. 2(c) memperlihatkan spektrum garis hasil modulasi frekuensi yang diusulkan Chowing.

Pengembangan untuk memperkaya pola sebaran spektrum pada teknik sintesis FM dikerjakan dengan beragam teknik [19]-[24]. Untuk keperluan penyandi yang diusulkan, dipilih teknik sintesis FM dari Chowing karena metode sintesis FM klasik ini membutuhkan waktu komputasi paling cepat dan paling efisien dalam merepresentasikan komponen harmonik.

II. METODOLOGI

Penyandian isyarat musik berbasis modulasi frekuensi dikerjakan dalam beberapa tahapan. Terdapat empat tahapan pokok yang harus dikerjakan. Secara garis besar estimasi parameter FM dapat digambarkan diagram blok seperti dalam Gbr. 3.

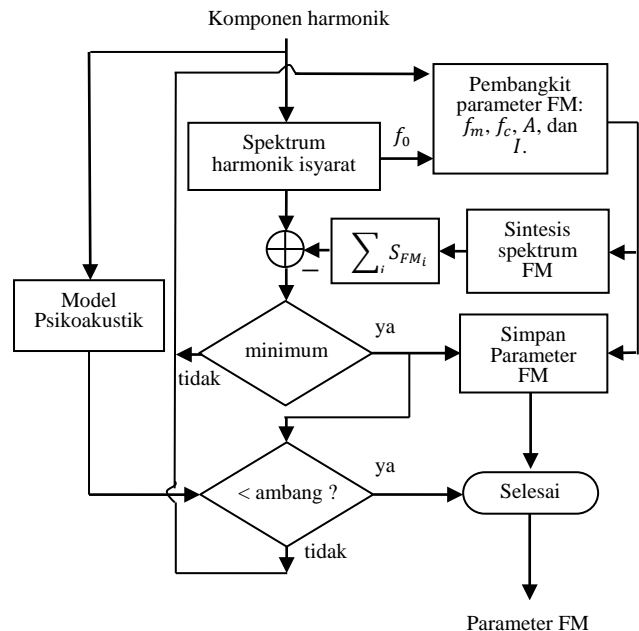


Gbr. 3 Prosedur penyandian isyarat musik berbasis modulasi frekuensi.

Langkah pertama adalah memotong isyarat *input* dengan jendela kosinus sepanjang 32 ms atau setara dengan 256 cuplikan. Jendela kosinus diletakkan secara *overlapped* setengah panjang jendela untuk mengurangi *block effect* akibat proses pemotongan isyarat. Dari setiap *frame* isyarat musik selanjutnya diekstrak komponen sinusoid [15].

Langkah kedua memisahkan komponen harmonik dengan komponen inharmonik berdasar pada [16]. Dalam ekstraksi komponen harmonik diasumsikan bahwa selalu hanya terdapat satu sumber harmonik. Seandainya dalam satu *frame* terdapat lebih dari satu sumber harmonik, tetap diasumsikan bahwa hanya terdapat satu perangkat sumber harmonik saja. Asumsi ini dipakai untuk menyederhakan kompleksitas analisis isyarat dan estimasi parameter FM. Hasil dari tahapan ini berupa komponen harmonik isyarat.

Langkah ketiga merupakan proses estimasi parameter FM berdasar pada [25]. Strategi yang dibuat untuk keberhasilan estimasi parameter FM yang diusulkan adalah berdasar pada hasil kompromi dari dua realita alami yang bersifat bertolak belakang. Di satu sisi, pola spektrum FM yang diproduksi dengan satu perangkat parameter FM mempunyai keterbatasan variasinya, dan di sisi lain, spektrum isyarat musik mempunyai variasi pola spektrum yang tidak terbatas, sehingga secara umum untuk merepresentasikan spektrum isyarat musik diperlukan lebih dari satu operator FM atau seperangkat parameter FM. Untuk menghemat komputasi, operator FM yang satu dengan operator FM yang lain diestimasi secara sekuensial dan dikerjakan dengan metode *analysis by sintesis*. Proses estimasi parameter FM dideskripsikan oleh diagram alir pada Gbr. 4.



Gbr. 4 Estimator parameter FM.

Frekuensi dasar f_0 dari komponen harmonik dipakai untuk menentukan frekuensi modulasi f_m dan frekuensi pembawa f_c . Karena teknik modulasi frekuensi akan dipekerjakan sebagai sumber spektrum harmonis, maka perbandingan frekuensi pembawa dan frekuensi modulasi berupa bilangan bulat. Hasil

pembangkitan parameter FM digunakan untuk mensintesis spektrum FM. Pada awal estimasi, spektrum hasil sintesis langsung dipakai untuk mengurangi spektrum harmonik asli. Parameter FM diatur secara iteratif dan dipilih untuk nilai selisih paling minimum. Seandainya selisih spektrum belum berada di bawah nilai ambang yang telah ditentukan, parameter FM berikutnya akan diestimasi dengan cara yang sama, hanya saja spektrum pengurang yang dipakai sekarang berasal dari jumlahan hasil sintesis spektrum FM yang baru ditambah dengan spektrum FM hasil sintesis parameter FM yang telah dipilih sebelumnya. Proses estimasi parameter FM ini berulang secara sekuensial sampai sisa selisih setiap magnitude komponen spektrum berada di bawah nilai ambang yang telah ditentukan oleh kurva *signal to masking ratio* (SMR) dari model psikoakustik untuk setiap *frame*.

Model psikoakustik merupakan model persepsi sistem pendengaran manusia. Model psikoakustik ini menghitung dan menentukan kurva *masking threshold* global akibat sifat *masking effect* dari tanggapan persepsi telinga terhadap suara. Pada penyandi yang dibuat, *masking threshold* digunakan untuk menyembunyikan kesalahan pemodelan FM isyarat audio sehingga tidak terdeteksi oleh persepsi pendengaran manusia. Kesalahan pemodelan disembunyikan di bawah kurva *masking threshold*.

Terdapat dua jenis model psikoakustik yaitu model psikoakustik 1 dan model psikoakustik 2. Hasil kedua model tidak berbeda jauh, tetapi model psikoakustik 1 memerlukan komputasi yang lebih ringan dibanding model psikoakustik 2 [26]-[27]. Berdasarkan pertimbangan tersebut, model psikoakustik 1 yang dipakai dalam pembuatan penyandi persepsi berdasar FM.

Langkah terakhir adalah proses penyandian parameter FM menjadi aliran bit. Kuantisasi yang dikerjakan dalam penyandian adalah kuantisasi skalar dengan lebar kuantisasi yang seragam maupun tidak seragam, tergantung dari probabilitas nilai parameter FM. Tabel I memuat jenis dan resolusi kuantisasi yang dipakai dalam penyandian.

TABEL I
KUANTISASI PARAMETER

No	Parameter	Kuantisasi (bit)	Jenis
1	Frekuensi Modulasi	13	logaritmis
2	Amplitude	8	logaritmis
3	Frekuensi Pembawa	5	linear
4	Indeks Modulasi	4	linear
5	Fase harmonik	5	linier
6	Prediktor	4	linier

Parameter 1 sampai 4 merupakan parameter FM. Parameter 5 adalah fase dari setiap komponen harmonik. Parameter fase ini diperlukan untuk rekonstruksi isyarat di bagian dekoeder atau penerima.

Pada sisi dekoder atau penerima tidak terjadi proses yang serumit di pengirim. Langkah proses rekonstruksi isyarat dari aliran bit yang diterima adalah sebagai berikut.

- Mengubah aliran bit menjadi parameter FM kembali. Proses dekuantisasi tidak akan mengembalikan parameter FM persis seperti hasil estimasi di pengirim.

- Mengubah parameter FM ke komponen harmonik isyarat dengan pertolongan fungsi Bessel jenis pertama. Jika dalam satu *frame* terdapat lebih dari satu operator FM, maka operator-operator FM disusun secara paralel. Komponen harmonik hasil rekonstruksi merupakan jumlah hasil sintesis komponen harmonik dari masing-masing operator FM.
- Rekonstruksi isyarat musik dari komponen harmonik.

III. HASIL DAN PEMBAHASAN

Isyarat musik yang dipakai dalam simulasi berasal dari instrumen musik tunggal seperti Klarinet, Harp, Trumpet dan Horn, dan isyarat musik dalam *ensemble* seperti Haendel dan Gamelan. Keseluruhan isyarat dicuplik dengan frekuensi cuplik 8 kHz dan dikuantisasi dengan resolusi 16 bit/cuplikan. Evaluasi tingkat keberhasilan penelitian dilakukan dengan membandingkan isyarat asli dengan isyarat hasil rekonstruksi pada ranah frekuensi dan waktu. Di ranah frekuensi isyarat ditampilkan dalam spektrum dan di ranah waktu dipresentasikan dengan betuk gelombang dan perbedaan kuantitatif dinyatakan dalam MSE.

A. Hasil Penggunaan Model Psikoakustik

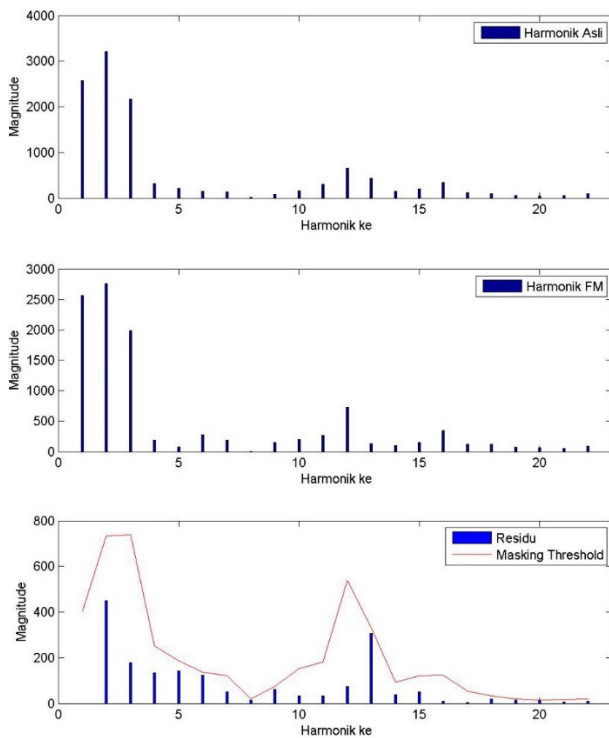
Seperti yang telah diberitakan pada [25], terdapat permasalahan dalam menentukan batas kesalahan pemodelan yang memenuhi kompromi antara minimal kebutuhan parameter model dan tingkat kesempurnaan model. Permasalahan penentuan batas kesalahan model pada [25] telah diselesaikan dengan pertolongan dari pemakaian sifat persepsi sistem pendengaran manusia. Sifat persepsi dalam diemuliskan dengan model psikoakustik dalam bentuk *masking threshold*.

Pada Tabel II dapat dilihat hasil pemodelan isyarat berbasis modulasi frekuensi diukur berdasarkan kebutuhan jumlah operator dan nilai MSE. Dibandingkan dengan hasil pada [25], jumlah kebutuhan operator FM dan nilai MSE berada diantara nilai ambang 1 dan 0,01. Hal ini menunjukkan bahwa pemakaian nilai ambang berdasar *masking threshold* merupakan kompromi yang optimal tanpa mengurangi kualitas secara drastis.

TABEL II
JUMLAH OPERATOR FM DAN NILAI MSE UNTUK NILAI AMBANG DARI MASKING THRESHOLD

Isyarat	Nilai ambang berdasar model psikoakustik	
	Jumlah Operator FM per Frame	MSE
Klarinet	3,4083	0.0012
Horn	5,0887	3.1696e-04
Trumpet	3,9041	1.2801e-04
Harp	3,8819	3.8459e-05
Haendel	7,9493	5.7618e-05
Gamelan	6,6776	3.8913e-04
Puspa8k	5,9464	0.0016
Kg	8,2009	0.0012
Castanets	9,2811	4.2108e-06
Speechm1	6,9620	2.0955e-04
Speechf1	7,2754	1.6726e-04
Female	6,8161	1.5006e-04

Pengujian pemakaian *masking threshold* juga dilakukan di ranah frekuensi. Gbr. 5 menunjukkan tiga spektral isyarat. Gbr. 5 sebelah atas adalah spektrum isyarat harmonik asli, Gbr. 5 bagian tengah adalah spektrum isyarat harmonis dari model FM, sedangkan Gbr. 5 paling bawah menunjukkan seluruh kesalahan pemodelan terletak di bawah kurva *masking threshold* (warna merah) pada akhir pemodelan.



Gbr. 5 Spektrum isyarat hasil rekonstruksi (atas), spektrum isyarat harmonis yang dimodelkan (tengah), dan spektrum residu yang berada dibawah *masking threshold*.

B. Hasil Penggunaan Prediktor

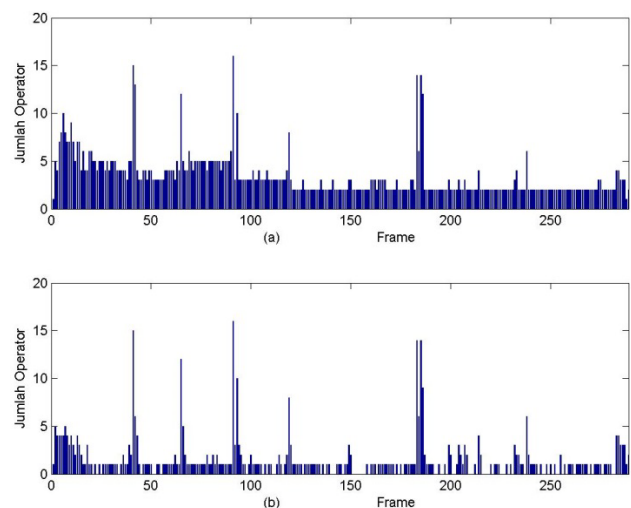
Isyarat musik tidak selamanya mempunyai dinamika frekuensi yang tinggi dan terus menerus di sepanjang sumbu waktu. Ada kalanya isyarat musik terasa stasioner di suatu nada frekuensi tertentu. Pada kasus seperti ini spektrum isyarat musik tidak banyak berubah dari *frame-frame* yang berturutan. Untuk mengurangi jumlah redundansi penggunaan operator pada *frame-frame* yang mempunyai kemiripan spektrum, sebuah prediktor dipekerjakan. Prediktor bekerja pada saat sebuah *frame* mempunyai jumlah harmonik yang sama dengan jumlah harmonik *frame* sebelumnya. Hasil simulasi pengaruh pemakaian prediktor dapat dilihat pada Tabel III.

Jumlah operator dapat berkurang sampai 61 % pada isyarat Harp. Menurunnya jumlah pemakaian operator secara dramatis menunjukkan bahwa isyarat Harp mempunyai jumlah harmonik yang relatif konstan antar *frame* dalam kurun waktu tertentu. Isyarat Klarinet menempati urutan kedua dari bawah dalam dinamika spektrum. Penghematan jumlah pemakaian operator dari *frame* ke *frame* dapat dilihat pada Gbr. 6. Akibat baik penggunaan prediktor tampak pada beberapa *frame* yang tidak perlu lagi memerlukan operator untuk mendeskripsikan komponen harmonik. Keseluruhan harmonik cukup diwakili

dengan sebuah koefisien prediktor. Meskipun demikian, tidak selamanya penggunaan prediktor membawa keuntungan yang tinggi. Pada isyarat dengan dinamika spektrum yang tinggi, seperti Speechm1, penghematan pemakaian operator FM hanya sebesar 0,61 %. Pada situasi seperti ini, penggunaan prediktor sebaiknya ditinjau kembali, karena penghematan 0,61 % tidak dapat membayar kembali biaya alokasi bit yang akan dipakai prediktor. Tetapi karena penghematan yang minimal sekali seperti pada isyarat Speechm1 adalah sangat jarang terjadi, maka penggunaan prediktor pada makalah ini tetap dikerjakan pada semua isyarat. Pengaruh pemakaian prediktor untuk isyarat Speechm1 dapat dilihat pada Gbr. 7.

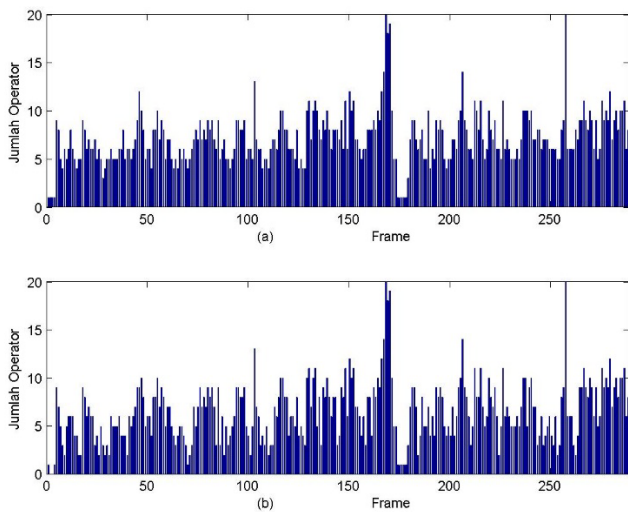
TABEL III
HASIL ESTIMASI PARAMETER FM TANPA DAN DENGAN PREDIKTOR

Isyarat	Rerata Operator FM / <i>Frame</i>		Reduksi operator akibat prediktor (%)
	Tanpa Prediktor	Dengan Prediktor	
Klarinet	3,4083	1,4775	56,6497
Horn	5,0887	2,2624	55,5394
Trumpet	3,9041	2,6232	32,8070
Harp	3,8819	1,4760	61,9772
Haendel	7,9493	7,0308	11,5544
Gamelan	6,6776	5,7976	13,1788
Puspa8k	5,9464	4,3496	26,8538
Kg	8,2009	5,5165	32,7335
Speechm1	9,2811	9,2242	0,61349
Speechf1	7,6827	6,9620	9,3806
Female	7,2754	6,6393	8,7426



Gbr. 6 Reduksi jumlah operator akibat penggunaan prediktor pada Klarinet.

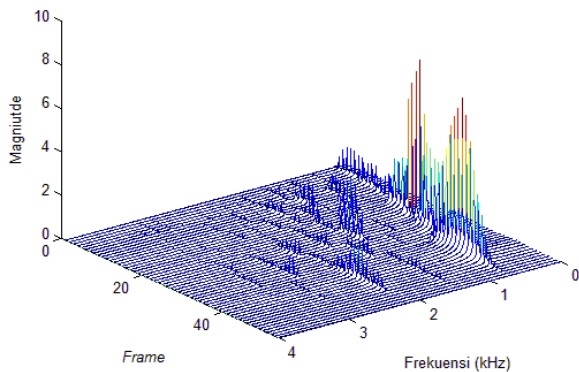
Isyarat Klarinet memerlukan jumlah parameter FM paling sedikit. Hal ini menunjukkan bahwa kompleksitas struktur harmonis isyarat Klarinet lebih sederhana dibanding kedua isyarat tutur. Isyarat tutur pria Speechm1 membutuhkan jumlah parameter FM yang lebih besar dibanding isyarat tutur wanita Speechf1. Fenomena ini menunjukkan jumlah harmonik pada isyarat tutur pria lebih tinggi. Penyebab tingginya jumlah harmonik disebabkan karena isyarat tutur pria mempunyai frekuensi dasar yang lebih rendah dibanding isyarat tutur wanita.



Gbr. 7 Reduksi minimal jumlah operator akibat penggunaan prediktor pada Speechm1.

C. Hasil Pengamatan di Ranah Waktu-Frekuensi

Selanjutnya simulasi dilaksanakan di ranah waktu-frekuensi. Gbr. 8 menunjukkan isyarat Klarinet dan Gbr. 9 memperlihatkan isyarat Klarinet hasil rekonstruksi dari parameter FM. Secara sepintas keduanya tidak mempunyai perbedaan yang jelas, tetapi kalau diamati lebih detail, isyarat asli mempunyai jumlah komponen frekuensi dengan magnitudo kecil yang lebih banyak. Komponen-komponen ini mempunyai sifat yang inharmonis, sehingga komponen ini tidak muncul lagi pada isyarat hasil rekonstruksi pada Gbr. 9.

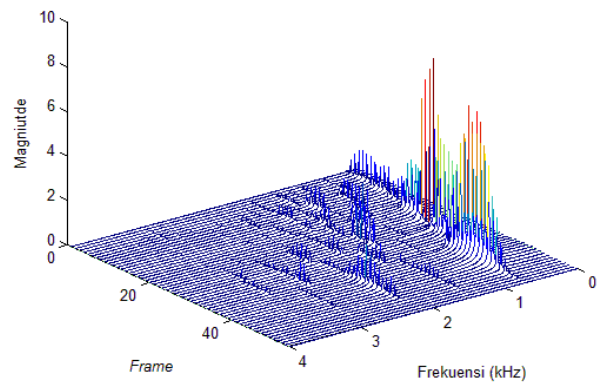


Gbr. 8 Isyarat Klarinet asli.

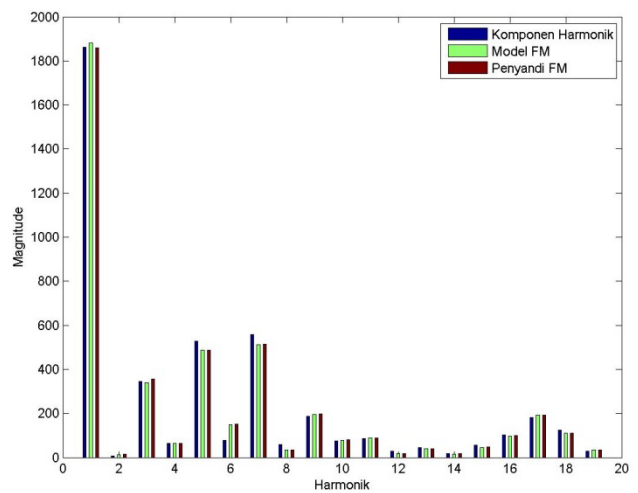
D. Pengamatan di Ranah Frekuensi

Pengujian di ranah frekuensi dilakukan dengan membandingkan spektrum komponen harmonik, isyarat hasil rekonstruksi dari model FM, dan isyarat hasil rekonstruksi pada penyandi FM. Gbr. 10 merupakan hasil pengamatan isyarat Klarinet pada *frame* ke-30. Proses pemodelan dan penyandian berbasis FM tidak memberikan selisih perbedaan yang signifikan terhadap komponen harmonik sebagai isyarat aslinya. Hal ini juga dibuktikan dari kualitas suara yang tidak mengalami penurunan yang signifikan, terutama pada isyarat Klarinet, Horn, dan Trumpet. Proses kuantisasi sendiri seharusnya menghilangkan sebagian komponen isyarat, tetapi

pada kenyataannya kadang bahkan mengembalikan komponen isyarat yang hilang pada saat proses pemodelan FM. Hal ini dapat dilihat pada harmonik ke-1.



Gbr. 9 Isyarat Klarinet hasil rekonstruksi dari parameter FM.



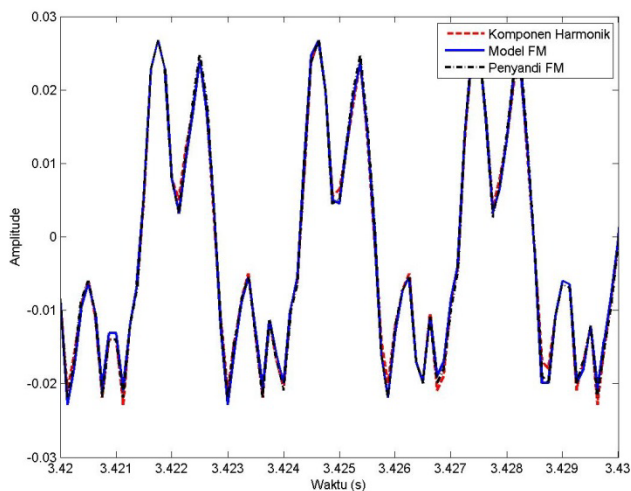
Gbr. 10 Perbandingan spektrum original komponen harmonik isyarat Klarinet *frame* ke-30, spektrum model FM, dan spektrum isyarat yang disintesis pada penerima.

E. Pengamatan di Ranah Waktu

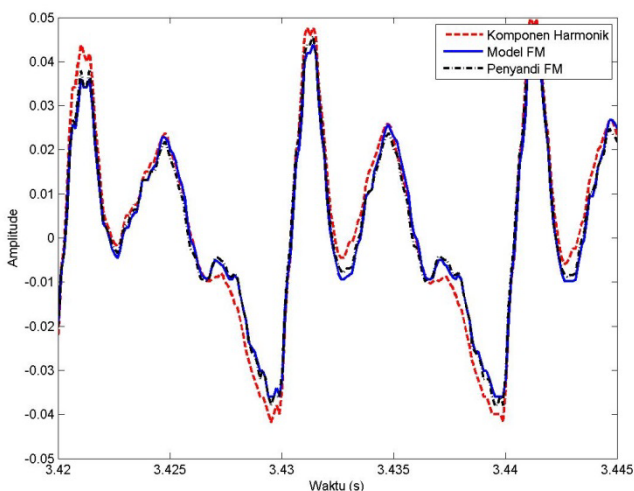
Pengujian di ranah waktu dilaksanakan dengan mengamati bentuk gelombang isyarat komponen harmonik, isyarat hasil rekonstruksi dari model FM, dan isyarat hasil rekonstruksi pada penyandi FM. Gbr. 11 merupakan hasil pengamatan isyarat Harp dalam lingkup waktu 3,420 ms sampai 3,430 ms. Proses kuantisasi tidak memberikan perbedaan yang jelas. Dengan demikian dapat disimpulkan bahwa resolusi kuantisasi untuk parameter FM sudah mencukupi.

Untuk memberikan pengamatan pembanding, penyandi dipekerjakan juga untuk isyarat tutur Speechm1. Hasil pengamatan di ranah waktu untuk isyarat tutur Speechm1 dapat dilihat pada Gbr. 12. Isyarat hasil sintesis dan penyandian tidak berbeda banyak. Hal ini menunjukkan resolusi kuantisasi yang digunakan sudah cukup halus. Jika dibanding dengan isyarat harmonik asli tampak adanya perbedaan yang relatif signifikan. Realita ini memperlihatkan bahwa pemodelan FM yang dikerjakan di sini kurang begitu cocok untuk isyarat

suara. Hal ini dapat diakibatkan karena dinamika frekuensi isyarat suara lebih tinggi dibanding isyarat Harp atau isyarat musik lainnya seperti Klarinet, Trumpet, dan Horn.



Gbr. 11 Perbandingan bentuk gelombang komponen harmonik isyarat Harp, gelombang model FM, dan gelombang isyarat yang disintesis pada penerima.



Gbr. 12 Perbandingan bentuk gelombang komponen harmonik isyarat Speechm1, gelombang model FM dan gelombang isyarat yang disintesis pada penerima.

F. Pengamatan Pesat Bit

Pengujian pesat bit dikerjakan dengan mengukur pesat bit dari masing-masing isyarat. Konfigurasi penyandi memakai teknik modulasi frekuensi klasik. Tabel IV menunjukkan hasil pengukuran pesat bit.

Pesat bit yang dihasilkan penyandi sangat tergantung pada jenis isyarat. Pesat bit rendah dicapai untuk isyarat musik instrumen tunggal dan mempunyai corak *partials* yang harmonis. Isyarat-isyarat tersebut dihasilkan oleh instrumen tiup seperti Klarinet, Horn, dan Trumpet. Untuk isyarat yang dihasilkan oleh instrumen *ensemble*, Haendel, Gamelan, Puspa8k, dan Kg. Meskipun kebutuhan pesat bit meningkat seiring dengan bertambahnya jumlah operator FM, tetapi laju kebutuhan bit tidak selalu sebanding dengan laju pertambahan

kebutuhan operator semata. Kenaikan pesat bit masih juga ditentukan oleh kompleksitas komponen harmonik karena parameter fase setiap komponen harmonik harus disandikan sendiri. Secara perkiraan kasar, kebutuhan bit untuk parameter fase dapat mencapai 40% kebutuhan bit keseluruhan.

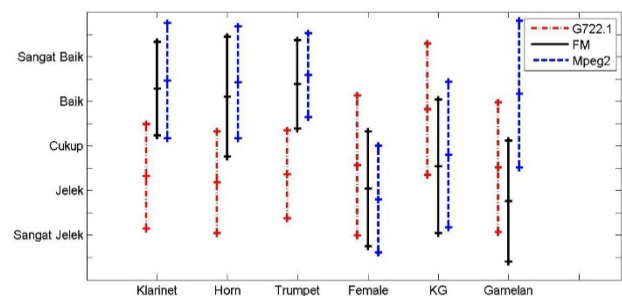
TABEL IV
KEBUTUHAN RERATA OPERATOR FM PER FRAME DAN PESAT BIT

Isyarat	Rerata Operator FM / Frame	Pesat bit
Klarinet	1.4671	5,0923 kbit/s
Horn	2.1534	6,5518 kbit/s
Trumpet	2.3356	6,4335 kbit/s
Harp	1.5092	6,0827 kbit/s
Haendel	7.0264	17,210 kbit/s
Gamelan	5.7903	15,076 kbit/s
Puspa8k	5.5102	14,325 kbit/s
Kg	4.3543	11,778 kbit/s
Speechm1	6.8241	16,120 kbit/s
Speechf1	6.5508	16,227 kbit/s
Female	5.4632	15,694 kbit/s

G. Hasil Tes Subyektif

Tes subyektif dilakukan dengan metode *Most Opinion Score* (MOS). Peserta tes pengujian yang berjumlah 35 orang diberi kesempatan untuk mendengarkan empat buah isyarat untuk setiap isyarat yang diujikan. Empat isyarat tersebut adalah isyarat asli dan isyarat hasil penyandian dari tiga metode yang dibandingkan, yaitu ITU G722.1, MPEG2 dan FM yang diusulkan. Tugas peserta adalah mendengarkan isyarat asli dan isyarat hasil penyandian dengan tiga metode yang berbeda. Selanjutnya peserta menilai kualitas masing-masing hasil penyandian berdasar kemiripannya dengan isyarat asli. Penilaian dilakukan dengan nilai 1, 2, 3, 4, dan 5 untuk kualitas sangat jelek, jelek, cukup, baik, dan sangat baik.

Dari keseluruhan isyarat yang dipakai untuk simulasi, hanya dipilih enam buah isyarat untuk pelaksanaan tes subyektif, yaitu Klarinet, Horn, Trumpet, Female, KG, dan Gamelan. Penyandi MPEG2 dan ITU-G722.1 diatur untuk bekerja pada pesat bit 16 kbit/s, sedangkan penyandi FM bekerja tanpa kekangan pesat bit. Untuk mendapat konsistensi penilaian, keenam isyarat diulang untuk didengarkan dengan urutan yang berbeda. Hasil tes subyektif ditampilkan dengan *confidence interval* 95% pada Gbr. 13.



Gbr. 13 Hasil tes subyektif penyandi ITU-G722.1, MPEG2 dan FM.

Untuk isyarat musik instrument tunggal, Klarinet, Horn, dan Trumpet, penyandi FM bekerja pada pesat bit 6 kbit/s mampu meraih kualitas MPEG2 dan jauh lebih baik dibanding ITU-G722.1, yang keduanya bekerja pada 16 kbit/s. Untuk isyarat tutur Female dan isyarat KG (vokal penyanyi), ketiga penyandi bekerja pada kualitas yang seimbang, meskipun penyandi FM berkerja pada pesat bit 11 kbit/s. Penyandi FM menorehkan kualitas yang terendah pada isyarat Gamelan. Penyandi FM tidak dapat bekerja optimal untuk isyarat gamelan disebabkan oleh sifat isyarat Gamelan yang tidak memenuhi asumsi yang mendasari pembuatan penyandi FM. Isyarat gamelan terbentuk oleh mayoritas instrumen musik jenis pukul. Proses pemukulan pada instrumen gamelan menghasilkan isyarat yang in-harmonik.

IV. KESIMPULAN

Sebuah penyandi persepsi isyarat audio berdasar pada model modulasi frekuensi telah berhasil dibuat. Dari hasil-hasil di tahapan proses penyandian dapat disimpulkan sebagai berikut.

Sebuah pemodelan isyarat musik berbasis pada teknik modulasi frekuensi di ranah frekuensi telah berhasil dikerjakan dengan baik. Replikasi spektrum isyarat musik di ranah frekuensi untuk setiap *frame* waktu memberikan hasil pemodelan yang akurat di ranah frekuensi dan waktu. Jumlah parameter FM yang diperlukan dalam pemodelan berbanding lurus dengan tingkat kompleksitas struktur harmonik isyarat dan tingkat kesempurnaan pemodelan.

Masalah penentuan batas kesempurnaan pemodelan isyarat berbasis modulasi frekuensi dapat diselesaikan dengan sangat baik dengan pemakaian model psikoakustik. Nilai ambang kesalahan untuk setiap garis spektrum ditentukan sesuai pengaruh garis-garis spektrum di sekitarnya dalam persepsi sistem pendengaran manusia.

Sebuah prediktor untuk mengeskloitasi redundans antar *frame* telah digunakan. Pemakaian prediktor dapat mereduksi jumlah operator FM sampai sekitar 60%.

Hasil tes subyektif dan uji pesat bit membuktikan dua fakta penting. Fakta pertama, penyandi FM mempunyai kinerja yang terbaik di antara penyandi MPEG2 dan ITU-G722.1 untuk isyarat musik tunggal. Hal ini dapat dilihat pada pesat bit di sekitar 6 kbit/s penyandi FM mempunyai kualitas setara dengan MPEG2 dan lebih baik dari penyandi ITU-G722.1, yang keduanya bekerja pada pesat bit 16 kbit/s. Fakta kedua, untuk isyarat musik *ensemble* dan isyarat tutur kinerja penyandi FM tidak lebih baik dibanding penyandi ITU-G722.1 dan MPEG2, sehingga dapat digarisbawahi bahwa penyandi FM ideal digunakan untuk isyarat musik instrumen tunggal.

REFERENSI

- [1] B. S. Atal, "Predictive coding of speech at low bit rates", *IEEE Transactions on Communications*, 30, P. 600-614, April 1982.
- [2] ITU-T. R. G.726., 40, 32, 24, 16 kbit/s Adaptive differential pulse code modulation (ADPCM), 1992.
- [3] ITU-T. R. G.729, Coding of speech at 8 kbit/s using conjugate-structure algebraic code excited linear-prediction (CS-ACELP), 1995.
- [4] ETSI Rec. GSM 06.10: GSM full rate speech transcoding, 1988.
- [5] I.-T. R. G.722, 7kHz audio coding within 64 kbit/s., vol. Fascicle III.4, Blue Book, P. 269-341, 1988.
- [6] I. JTC1/SC29/WG11, Coding of moving pictures and audio MPEG-2 Advanced Audio Coding, ISO/IEC 13818-7 international standard, 1997.
- [7] I. JTC1/SC29/WG11, Coding of moving pictures and audio MPEG-2 Advanced Audio Coding, ISO/IEC 13818-7 international standard, 1997.
- [8] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding", *J. Audio Eng. Soc.*, 1997.
- [9] I. JTC1/SC29, Mpeg4 Audio coding, ISO/IEC 14496-3:/Amd.1:1999(E),, 1999.
- [10] S. Quackenbush, "MPEG Unified Speech and Audio Coding", *IEEE Multimedia*, vol. 20, no. 2, pp. 72-78, April-June 2013.
- [11] B. Winduratna, A. Susanto dan R. Hidayat, "Penyandi isyarat Musik pada Pesat Bit di bawah 16 kbit/s Berdasar Model Isyarat Modulasi frekuensi (FM)", Disertasi, Departemen Teknik Elektro dan Teknologi Informasi, FT UGM, Yogyakarta, 2016
- [12] G. Eska, "Schall und Klang: Wie und was wir hören", Birkhäuser Verlag, 1997.
- [13] M. Goodwin, *Adaptive Signal Models, Theory, Algorithms and Audio Application*, Kluwer Academic Publishers, Netherland, 1998.
- [14] S. Kay, "A Fast and Accurate Single Frequency Estimator, IEEE", *Transactions on Acoustics, Speech, Signal Processing*, Vol. 37, No. 12, p. 1987-1990., 1989.
- [15] B. Winduratna, A. Susanto dan R. Hidayat, "Ekstraksi Komponen Sinusoida dari Isyarat Audio", *Annual Engineering Seminar*, FT UGM, Februari 2014
- [16] B. Winduratna, A. Susanto dan R. Hidayat, "Ekstraksi Komponen Harmonik dari Isyarat Audio", *Annual Engineering Seminar*, FT UGM, Februari 2015
- [17] E. Amstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation", *Proceeding of the Institute of Radio Engineers*, vol. 24, number 5, pp. 689-740, May 1936.
- [18] J. Chowning, , "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation", *Journal of the Audio Engineering Society*, 21 (7), 1973
- [19] B. Schottstaedt, "The Simulation of Natural Instrument Tones Using Frequency Modulation with a Complex Modulating Wave", *Computer Music Journal*, Vol. 1, No. 4, pp. 46-50, 1977.
- [20] J.-P. Palamin, P. Palamin, A. Ronveaux, "A Method of Generation and Controlling Musical Asymmetrical Spectra", *Journal Audio Eng. Soc.*, Vol. 36, pp. 671-685, Sept. 1988.
- [21] B.T.G. Tan, S.L. Gan, "Real-Time Implementation of Double Frequency Modulation (DFM) Synthesis", *Journal Audio Eng. Soc.*, 1994.
- [22] P. E. Etchemendy, M. C. Eguia and B. Mész, "Principal pitch of frequency-modulated tones with asymmetrical modulation waveform: A comparison of model", *The Journal of the Acoustical Society of America*, 135, 2014.
- [23] B. T. G. Tan, and S. M. Lim, "Automated Parameter Optimisation for Double Frequency Modulation Synthesis Using the Genetic Annealing Algorithm", *In: Journal of Audio Engineering Society*, 44, (1/2) pp 3-15. 1996.
- [24] T. J. Mitchell, and C. W. Sullivan, "Frequency Modulation Tone Matching Using a Fuzzy Clustering Evolution", *In: Proceedings of the 118th Convention of the Audio Engineering Society*, Barcelona, Preprint 6366, 2005.
- [25] B. Winduratna, A. Susanto dan R. Hidayat, "Pemodelan Isyarat Musik Berbasis pada Modulasi frekuensi (FM)", *CITEE 2015*, DTETI, FT UGM, September 2015
- [26] Jayant, N., "Noll, P & Pan, D.: The psychoacoustic model," in *Signal Compression: Coding of Speech, Audio, Text, Image and Video*, *World Scientific*, 1997, pp. 99-105.
- [27] Y. Lin, and W. H. Abdulla, *Audio Watermark*, pp 15-49, Springer, 2015.