

Adaptif Audio Watermarking Berdasarkan Nilai Snr Pada File Audio Dengan Informasi Sisipan Teks

Redi Kuncoro Katri^{1,*}, Gelar Budiman¹, Ledy Novamizanti¹

¹ Universitas Telkom, Fakultas Teknik Elektro

* redikuncorokatri@gmail.com

Abstrak. Menyisipkan informasi kedalam file digital adalah hal yang penting di era sekarang ini. Era dimana persebaran informasi sangat mudah dilakukan. *Digital watermarking* dilakukan dengan tujuan untuk menyisipkan informasi pada file digital tanpa mengubah file tersebut secara signifikan. *Audio watermarking* adalah salah satunya. Pada umumnya audio watermarking dilakukan untuk menyisipkan identitas pemilik atau pencipta file audio tersebut. Dalam penelitian ini watermarking dilakukan pada file audio dengan sisipan berupa teks. Dilakukan dengan menggunakan metode penyesuaian intensitas sisipan terhadap audio host agar *Signal to Noise Ratio* (SNR) yang terbentuk akibat watermarking selalu konstan dan mampu diekstraksi informasi sisipannya tanpa harus menggunakan audio aslinya (*blind watermark*). Penyesuaian intensitas sisipan dilakukan dengan harapan sisipan tidak terlalu mengganggu host audio dan tidak terganggu host audio. Dengan sekema ini, intensitas sisipan di setiap region sisipan akan berbeda beda sesuai dengan energi audio pada region tersebut. *Discrete Wavelet Transform* (DWT) dilakukan dalam penelitian ini dengan maksud menyisipkan informasi sisipan kedalam frekuensi yang kurang begitu terdengar oleh telinga manusia sehingga watermarking menjadi lebih inaudible atau tidak terdengar. Hasil akhir dari penelitian terbentuk skema atau metode watermarking yang kuat dan tidak terdengar dengan *Bit Error Rate* (BER) hingga dibawah 2% dengan $35 \text{ dB} < \text{SNR} < 50 \text{ dB}$

Kata Kunci: Adaptif, *Audio watermarking*, DWT, perlindungan hak cipta, SNR

1. Pendahuluan

Teknologi di dunia ini tidak pernah jenuh untuk berkembang. Jarak dan waktu seolah kini tidak menjadi masalah ketika internet telah menjadi bagian dari hidup kita. Penyebaran informasi pun menjadi sangat mudah dan seolah tanpa hambatan serta batasan. Mengingat hal tersebut, maka perlu dibuat sebuah skema agar segala data yang tersebar ini memiliki informasi kepemilikan yang jelas termasuk pada data audio.

Audio watermarking adalah suatu teknik menyisipkan informasi dengan maksud tertentu kedalam host media berupa audio tanpa mengganggu kualitas audio aslinya [1]. *Watermarked audio* yang memuaskan memiliki tiga kriteria yaitu informasinya tidak terdengar oleh telinga manusia, informasinya tegap dan kuat, serta tahan terhadap berbagai gangguan sinyal seperti random cropping, scale modification dan lain sebagainya [2].

Penyisipan dengan intensitas yang proporsional pada area sisipan menjadi kuncinya. Daerah dengan infleksi tinggi pada umumnya merupakan daerah transisi musik seperti permainan piano pada orkestra, permainan gitar solo pada musik rock dan lain sebagainya atau daerah dengan instrumen perkusi seperti drum, bell, atau ledakan yang merepresentasikan tempo atau ritme dari musik itu sendiri, yang mana tempo atau ritme tersebut merupakan bagian penting dari audio bagi telinga manusia [3]. Sehingga intensitas yang cukup besar pada daerah tersebut membuat informasi tidak mudah terhapus. Sebab cropping atau signal processing lainnya pada daerah tersebut, membuat file audio akan terdengar berbeda dengan aslinya. Sementara intensitas informasi pada region yang memiliki koefisien rendah harus pula cukup rendah agar sisipan tidak terlalu mengganggu dan tidak terdengar oleh telinga manusia. Penelitian ini merancang sebuah sekema algoritma adaptive watermarking berdasarkan *signal to noise ratio* (SNR) untuk penentuan skala intensitas penyisipan informasi dalam region tertentu pada audio.

Metode transformasi wavelet diskrit sendiri merupakan transformasi sinyal yang hasil transformasinya mampu memberikan informasi dalam domain frekuensi dan sekaligus dalam domain waktu. Sehingga memudahkan dalam penyisipan pada waktu dan frekuensi tertentu dari suatu sinyal.

Watermarking menggunakan DWT (*Discrete Wavelet Transform*) akan menghasilkan watermark yang kuat dibanding dengan metode lain [2].

Untuk penentuan daerah watermarking pada audio, penelitian [1] [2] [3] mengusulkan untuk menentukannya berdasarkan daerah yang memiliki energi tinggi yang merepresentasikan music edges seperti transisi musik, suara drum, piano dan lain sebagainya yang merupakan bagian penting dari musik itu sendiri. Pengalokasian sisipan pada daerah tertentu seperti ini bertujuan untuk menguatkan dan mampu bertahan dari serangan berupa *time-scaling modification* (TSM) yang menjadi kendala serius dalam setiap metode watermarking. Selain penempatan sisipan informasi yang harus tepat, penelitian [2] [3] juga mengusulkan untuk melakukan penyesuaian intensitas sisipan berdasarkan nilai SNR yang dihitung. Ini membuat intensitas sisipan menjadi adaptif dan tidak seragam di setiap tempat. Pada penelitian-penelitian adaptif audio watermarking sebelumnya [2] [3] proses ekstraksi sisipan tidak bias dilakukan tanpa adanya audio asli yang belum diberi watermark.

Sehingga menyisipkan informasi pada domain wavelet dengan intensitas sisipan yang adaptif akan membuat sisipan kuat, tidak terdengar dan tahan terhadap serangan dengan proses ekstraksi tanpa audio asli.

2. Teori dan Tahap Perancangan

Audio watermarking adalah suatu teknik menyisipkan informasi dengan maksud tertentu kedalam host media tanpa mengganggu kualitas aslinya [1]. Pada dasarnya penyisipan informasi ini dilakukan dengan cara menjumlahkan host media dengan informasi dengan intensitas sisipan tertentu sedemikian sehingga SNR sisipan menjadi konstan dan disisipkan pada domain wavelet di koefisien detailnya.

2.1 Segmentasi Audio dan Discrete Wavelet Transform

Pada tahap ini file audio disegmentasi dengan ukuran 2^{16} . Ukuran ini telah memperhitungkan kapasitas teks sisipan yang akan disisipkan. Pada penelitian ini, teks sisipan yang akan disisipkan memiliki panjang maksimum adalah 400 karakter dengan masing-masing karakter disimbolkan dengan 8 bit biner. Sehingga panjang informasi sisipan maksimum adalah 3.200 bit. Ditambah dengan header sebanyak 50 bit, maka data sisipan menjadi 3.250 bit. Dari data sisipan ini, kemudian akan dilakukan spreading dengan kode unik sepanjang 8 bit. Panjang data yang akan disisipkan pun menjadi 3.250×8 atau 26.000 bit data. Karena data ini akan disisipkan kedalam koefisien detail DWT dari masing-masing segmen, maka ukuran koefisien detail ini haruslah lebih besar atau setidaknya sama dengan 26.000 data. Angka faktor pelipatan dua yang terdekat dan lebih besar dari 26.000 adalah 32.768 atau 2^{15} . Namun bukan 2^{15} ukuran frame yang kita pakai. Dalam proses DWT, ukuran koefisien detail DWT satu level dari suatu sinyal adalah setengah dari ukuran sinyal awalnya. Sehingga, agar 26.000 data tersebut dapat disisipkan dalam koefisien detail DWT, ukuran segmennya adalah 2^{16} data.

Dari ukuran frame tersebut diatas, maka ukuran file audio yang akan disisipkan harus lebih besar atau setidaknya sama besar ukurannya dengan ukuran satu segmen. Maka jika sample rate audio adalah 44100 sample per detik, panjang audio yang akan diberikan sisipan minimal adalah sepanjang satu setengah detik. Kemudian dilakukan DWT Haar untuk masing-masing segmen sehingga didapatkan komposisi sinyal segmen tersebut berupa koefisien detail sinyal dan koefisien aproximity sinyal.

Discrete Wavelet Transform (DWT) dilakukan dengan melewati sinyal yang akan di DWT kedalam HPF dan LPF. Masing masing filter akan menghasilkan keluaran sinyal yang kemudian dilakukan downsampling sehingga panjang masing-masing sinyal keluaran filter adalah sepanjang setengah dari sinyal sebelumnya. Proses ini juga sering disebut sebagai proses dekomposisi sinyal. Output LPF yang telah di-downsampling disebut sebagai koefisien aproximity dan keluaran dari HPF adalah koefisien detailnya. Semakin besar ukuran segmen ini maka jumlah informasi yang bisa disisipkan pun menjadi semakin banyak.

2.2 Penyesuaian Intensitas terhadap SNR

Dalam tahap ini sinyal data sisipan akan ditentukan nilai intensitas sisipannya. Dari yang tadinya biner diubah menjadi bernilai 1 atau -1 (*Not Return to Zero* (NRZ)) dan akan dibuat lebih kecil dengan penskalaan sebesar α untuk kemudian ditambahkan dengan sinyal audio yang akan disisipkan informasi.

Misal jika sebuah sinyal audio x memiliki panjang n data, dan akan disisipkan data v sepanjang m data NRZ dengan intensitas sisipan sebesar α , dan menjadi audio akhir berupa audio y , dengan panjang n data pula. maka nilai *Mean Square Error* (MSE) adalah:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y(i) - x(i))^2 \quad (1)$$

Dengan metode penyisipan untuk m data pertama adalah:

$$y(i) = x(i) + \alpha(v(i)) \quad (2)$$

Dan untuk i di luar m data pertama adalah :

$$y(i) = x(i) \quad (3)$$

Sehingga pada persamaan MSE, $y(i) - x(i)$ akan bernilai $\alpha(v(i))$. Sehingga :

$$MSE = \frac{1}{n} \sum_{i=1}^m (\alpha \times v(i))^2 \quad (4)$$

Karena data v berupa data NRZ, yaitu data dengan nilai $+1$ atau -1 , maka $(\alpha \times v(i))^2$ akan sama dengan α^2 . Sehingga MSE watermarking menjadi :

$$MSE = \frac{m(\alpha^2)}{n} \quad (5)$$

Dengan persamaan SNR :

$$SNR = 10 \log \left(\frac{\frac{1}{n} \sum_{i=1}^n x^2(i)}{MSE} \right) \quad (6)$$

Maka

$$10^{SNR/10} = \frac{\frac{1}{n} \sum_{i=1}^n x^2(i)}{\frac{m(\alpha^2)}{n}} \quad (7)$$

$$10^{SNR/10} = \frac{\sum_{i=1}^n x^2(i)}{m(\alpha^2)} \quad (8)$$

Sehingga untuk penentuan nilai α adalah :

$$\alpha = \sqrt{\frac{\sum_{i=1}^n x^2(i)}{m \times 10^{SNR/10}}} \quad (9)$$

Intensitas α ini lah yang akan digunakan pada penyisipan watermark.

2.3 Penyisipan

Data sisipan berupa teks, dikodekan terlebih dahulu menjadi bit stream $w(i)$ yang masing-masing karakter teks dikodekan dengan 8 bit berdasarkan nilai asciinya Data biner ini kemudian ditambahkan data header dibagian awal. Data header ini terdiri dari 32 bit biner dengan pola khusus yang disepakati di penyisipan dan pengestraksian dan 18 bit yang setiap 9 bitnya menunjukkan jumlah karakter dalam teks sisipan. Sehingga panjang bit sisipan adalah $8 \times$ jumlah karakter sisipan + 50 bit. Yang kemudian dijadikan NRZ (*Not Return to Zero*) yang artinya menjadi rangkaian data dengan nilai 1 atau -1. Data ini kemudian kita sebut sebagai data $v(i)$. $v(i)$ yang telah berupa data NRZ dilakukan spreading dengan kode 8 bit. Yaitu dengan mengganti nilai $v(i)$ dengan (kode 8 bit $\times v(i)$). Missal kode adalah $s(i)$ maka $v'(i)$ setelah spreading adalah

$$v'(i) = s(i) \times v((i + 1) \text{ div } 8) \quad (10)$$

Sehingga semisal data $v=\{1,-1,-1,1\}$ dengan kode $s=\{1,-1,1,1,-1,-1,1,1\}$ maka $v'=\{1,-1,1,1,-1,-1,1,1,-1,1,-1,1,1,-1,1,-1,1,1,-1,1,1,-1,1,1,1\}$.

Kemudian sisipan dilakukan dengan menambahkan data koefisien detail DWT setiap segmen audio dengan data v' ini.

$$y(i) = \begin{cases} x(i) + \alpha(v'(i)) & \text{untuk } 1 < i < \text{panjang } v' \\ x(i) & \text{untuk } i \text{ lainnya} \end{cases} \quad (11)$$

Dimana x adalah data koefisien DWT dan α adalah intensitas sisipan dan v' merupakan data informasi NRZ yang telah di-spreading dan menghasilkan koefisien DWT baru berupa data y .

2.4 Pengambilan kembali sisipan

Anggap sinyal watermarked yang akan dideteksi ini dalam notasi $y'(i)$ yang mana merupakan $y(i)$ yang telah mengalami gangguan. Sinyal $y'(i)$ kemudian dilakukan proses segmentasi. Kemudian lakukan DWT pada masing masing segmen, dan kemudian ambil datanya pada setiap segmen dan dirata-rata data semua segmen untuk diambil keputusan data informasi akhir dari watermark tersebut.

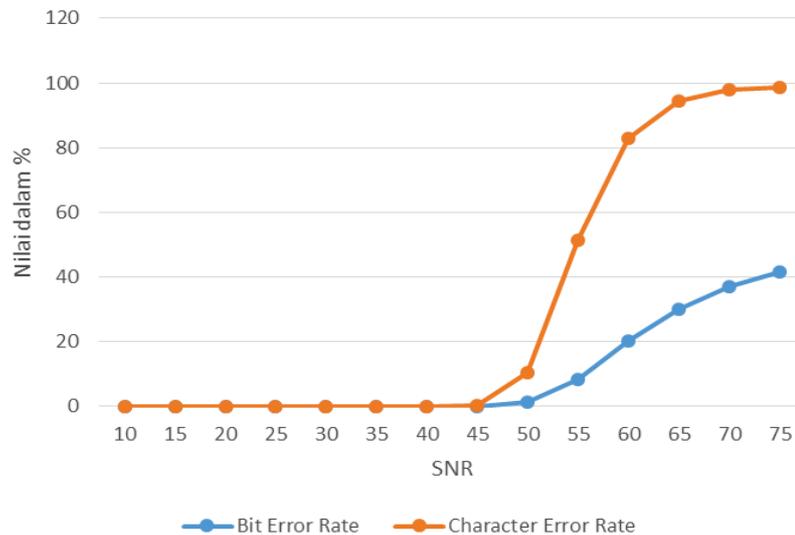
$$v(i) = \text{sign} \sum_{j=(i-1) \cdot 8 + 1}^{j \cdot 8} y'(j) \times s(j \bmod 8) \quad (12)$$

Dimana $s(i)$ adalah kode unik saat spreading data. Dan fungsi sign digunakan untuk mengambil nilai tanda hasil yang terbentuk. Apakah positif atau negative.

Hasil $v(i)$ ini kemudian dikembalikan dari data NRZ kedalam kode biner. Lalu diubah kedalam karakter sesuai ascii untuk setiap 8 kode biner yang terbentuk.

3. Hasil dan pengujian

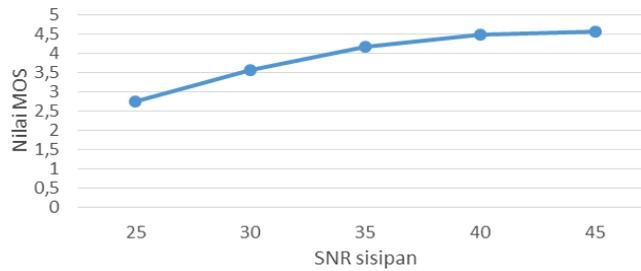
Hasil penyisipan informasi pada file audio dengan metode ini adalah sebagai berikut:



Grafik 1 SNR terhadap BER dan CER

Dari tabel diatas dapat dilihat bahwa semakin besar SNR target yang diinginkan pada penyisipan, BER dan CER pun semakin besar. Hal ini terjadi karena intensitas sisipan yang adaptif terhadap nilai SNR dan energi audio dan intensitas sisipan berbanding terbalik dengan nilai SNR yang dihasilkan seperti pada persamaan (7) dan (8). Sehingga semakin besar SNR maka sisipan akan semakin kecil dan susah untuk diekstraksi.

Sedangkan hasil pengaruh besar kecilnya SNR saat penyisipan terhadap *Mean Opinion Score* (MOS) adalah sebagai berikut :



Grafik 2 Pengaruh SNR terhadap MOS

Dengan melihat hasil MOS diatas, terlihat bahwa metode penyisipan adaptif ini sangat baik dan tidak begitu terdengar untuk sisipan yang SNR sisipan > 35.

Pengaruh ukuran frame terhadap kapasitas sisipan

Tabel 1 Pengaruh ukuran frame terhadap proses watermarking

Ukuran frame	10 karakter	50 karakter	100 karakter	200 karakter	300 karakter	400 karakter
2^{10}	Gagal	Gagal	Gagal	Gagal	Gagal	Gagal
2^{11}	Tersisipkan	Gagal	Gagal	Gagal	Gagal	Gagal
2^{12}	Tersisipkan	Gagal	Gagal	Gagal	Gagal	Gagal
2^{13}	Tersisipkan	Tersisipkan	Gagal	Gagal	Gagal	Gagal
2^{14}	Tersisipkan	Tersisipkan	Tersisipkan	Gagal	Gagal	Gagal
2^{15}	Tersisipkan	Tersisipkan	Tersisipkan	Tersisipkan	Gagal	Gagal
2^{16}	Tersisipkan	Tersisipkan	Tersisipkan	Tersisipkan	Tersisipkan	Tersisipkan

Dari tabel diatas dapat dilihat bahwa semain besar segmen, maka semakin besar kapasitas sisipan yang mampu disisipkan pada audio. 400 karakter adalah jumlah yang cukup untuk sebuah informasi kepemilikan sebuah audio. Sehingga ukuran segmen 2^{16} adalah ukuran segmen yang ideal.

4. Kesimpulan

Dilihat dari ketahanan informasi dan ketidakterdengarnya informasi dalam audio dapat disimpulkan bahwa dengan skema penyisipan adaptif ini, SNR sisipan yang paling maksimal adalah $35 \text{ dB} < \text{SNR} < 45 \text{ dB}$. Dengan nilai BER tanpa penyerangan sebesar 0% dan nilai MOS > 3,5.

Ukuran segmen 2^{16} merupakan ukuran yang cukup untuk memyisipkan watermark teks dengan kapasitas 400 karakter.

5. Daftar Referensi

- [1] Wei Li, Xiangyang Xue, and Peizhong Lu, "Localized Audio Watermarking Technique Robust Against Time-Scale Modification" IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 8, NO. 1, PP. 60-69, FEBRUARY 2006
- [2] Xueyao Li, Min Zhang, Shiliang Sun, "ADAPTIVE AUDIO WATERMARKING ALGORITHM BASED ON SNR IN WAVELET DOMAIN " IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2003, pp. 287-292
- [3] WU Guo-min (吴国民)†, ZHUANG Yue-ting (庄越挺)†‡, WU Fei (吴飞)†, PAN Yun-he (潘云鹤) , "Adaptive audio watermarking based on SNR in localized regions" Journal of Zhejiang University SCIENCE, 2005 6A(Suppl. D):53-57 *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.