

BAB IV ANALISIS DAN PEMBAHASAN

4.1. Analisis Kinerja

Penelitian dilakukan untuk menganalisis pengaruh sistem hierarki *downmix MPEG Surround* terhadap pesan rahasia yang disisipkan. Analisis kinerja difokuskan pada dua aspek utama, yaitu keberhasilan deteksi *watermark* dan *imperceptibility* (ketidakterlihatan/ketidakepaan) audio stego terhadap sinyal audio aslinya.

Ketahanan *watermark* dievaluasi berdasarkan perbandingan antara energi *watermark* yang disisipkan dengan nilai *error probability* hasil proses ekstraksi setelah melewati seluruh tahapan sistem. Semakin kecil nilai *error probability*, dimana dinyatakan menggunakan skala logaritmik (\log_{10}), untuk tingkat energi *watermark* tertentu, semakin baik kemampuan sistem dalam mempertahankan *watermark* terhadap proses transformasi audio. Sementara itu, *imperceptibility* dinilai melalui nilai *Signal-to-Noise Ratio* (SNR) dari sinyal audio yang telah melewati rangkaian proses penyisipan, *downmixing*, dan ekstraksi. Nilai SNR digunakan untuk melihat sejauh mana penyisipan *watermark* dan proses *MPEG Surround* memengaruhi kualitas sinyal audio akhir.

Parameter yang divariasikan pada percobaan adalah ukuran *frame* dan *embedding energy* (*tow u*). Ukuran *frame* yang digunakan adalah 1024, 2048, 3072, 4096, dan 5120. Sedangkan untuk *embedding energy* divariasikan dari 0.0001 hingga 0.9, sesuai dengan kebutuhan sistem untuk mencapai nilai *Error probability* yang maksimal pada tahap ekstraksi *watermark*. Begitu juga dengan pengujian nilai SNR, perlakuan yang sama dilakukan seperti pada pengujian ketahanan *watermark*, dimana nilai SNR diambil dari rata-rata seluruh variasi variabel penyisipan yang diuji.

4.2. Analisis *Watermark Energy* terhadap *Error probability*

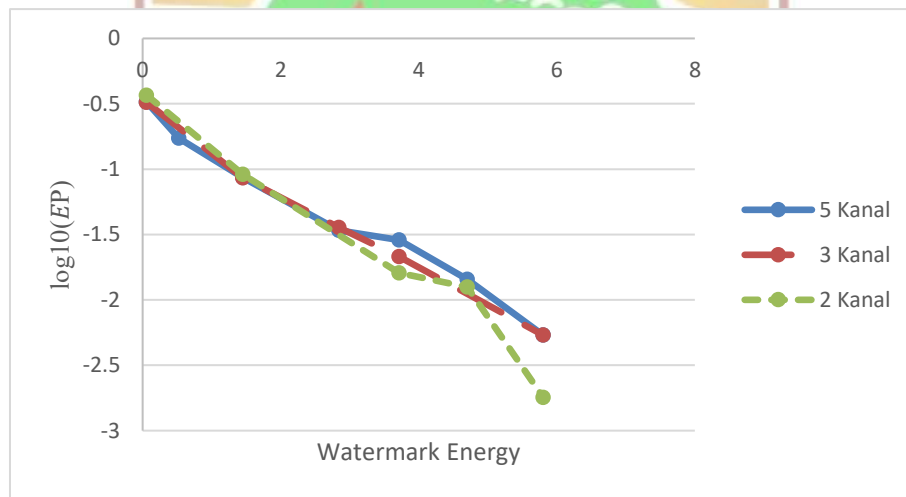
Analisis ini bertujuan untuk mengetahui sejauh mana peningkatan energi *watermark* dapat meningkatkan ketahanan informasi tersembunyi dari serangkaian proses sistem. Hubungan antara *watermark energy* dan *error probability* bersifat saling berkaitan. Secara umum, peningkatan *watermark energy* dapat memperkuat kejelasan sinyal *watermark* sehingga menurunkan probabilitas kesalahan yang terjadi pada proses ekstraksi *watermark*. Namun, hubungan ini tidak selalu linear, karena karakteristik sinyal audio pada setiap kanal, efek *downmixing*, dan proses kompresi juga memengaruhi hasilnya. Nilai EP yang dinyatakan pada grafik menggunakan skala logaritmik (\log_{10}).

4.2.1. Penyisipan sebelum *Downmix*

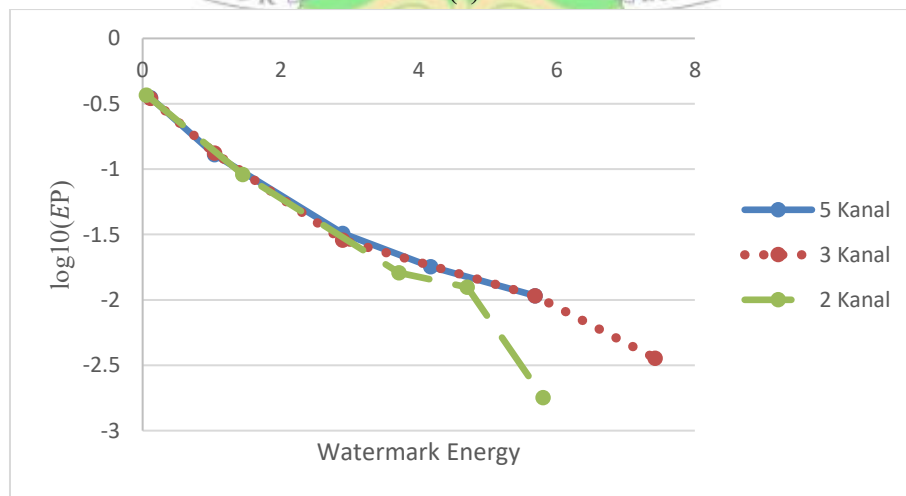
Pada tahapan proses penyisipan sebelum *downmix MPEG Surround*,

dilakukan eksperimen untuk menganalisis pengaruh pengurangan level hierarki *downmix* terhadap performa *watermarking*. Mekanisme penyisipan dapat dilihat berdasarkan pada gambar 3.2 (a) s.d. (c). Gambar 3.2 (a) menerapkan proses penyisipan pesan rahasia pada seluruh kanal audio sebelum proses *downmix*. Setiap kanal input masing-masing melalui proses penyisipan, setelah itu audio yang telah disisipi pesan rahasia mengikuti hierarki *downmix* utuh *MPEG Surround* yang akan menghasilkan output berupa mono. Mekanisme kedua, Gambar 3.2 (b) menerapkan penyisipan pada 3 kanal sebelum proses *downmix* dengan hierarki yang dikurangi 1 level. Begitu juga dengan Gambar 3.2 (c) yang menerapkan penyisipan 2 kanal sebelum proses *downmix* dengan hierarki yang dikurangi 2 level.

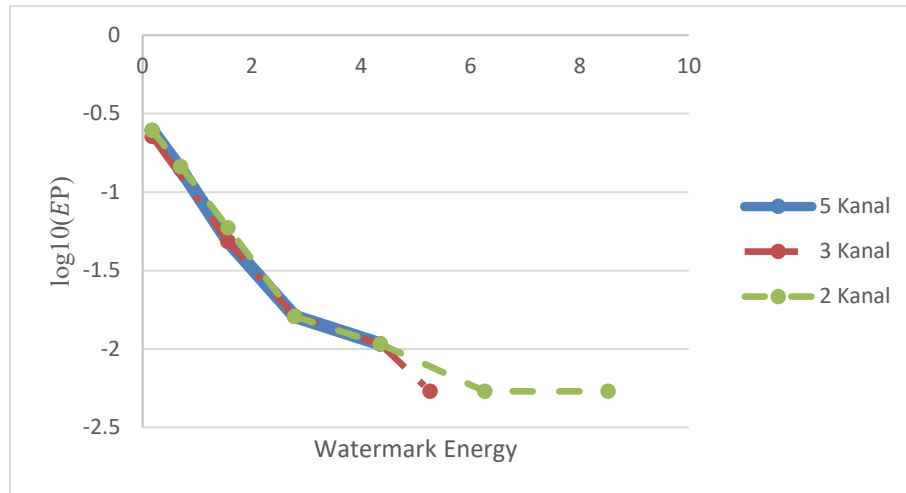
Ketiga mekanisme ini dirancang untuk memberikan perbandingan komprehensif mengenai bagaimana jumlah kanal yang disisipkan pesan rahasia dan kompleksitas hierarki *downmix* mempengaruhi ketahanan *watermark*. Evaluasi difokuskan pada hubungan antara *watermark energy* yang digunakan dengan *error probability* yang dihasilkan, bertujuan untuk menemukan konfigurasi optimal yang memberikan keseimbangan antara kekuatan *watermark* dan akurasi deteksi.



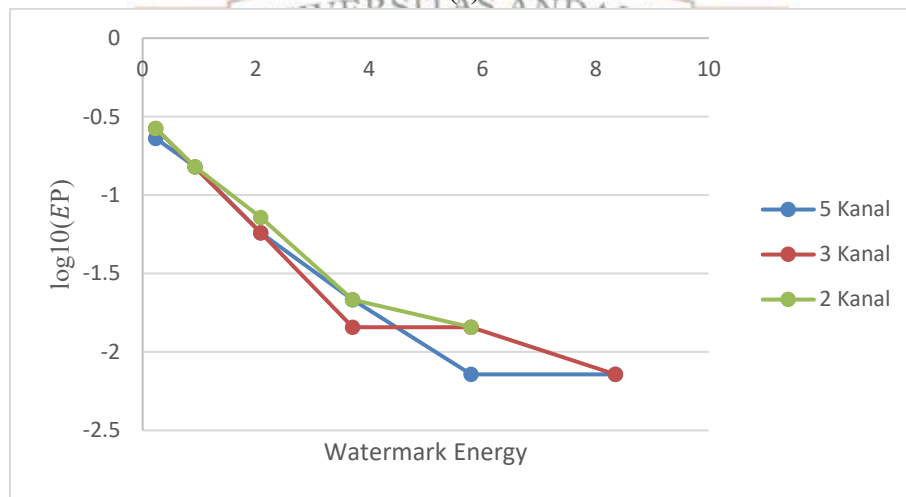
(a)



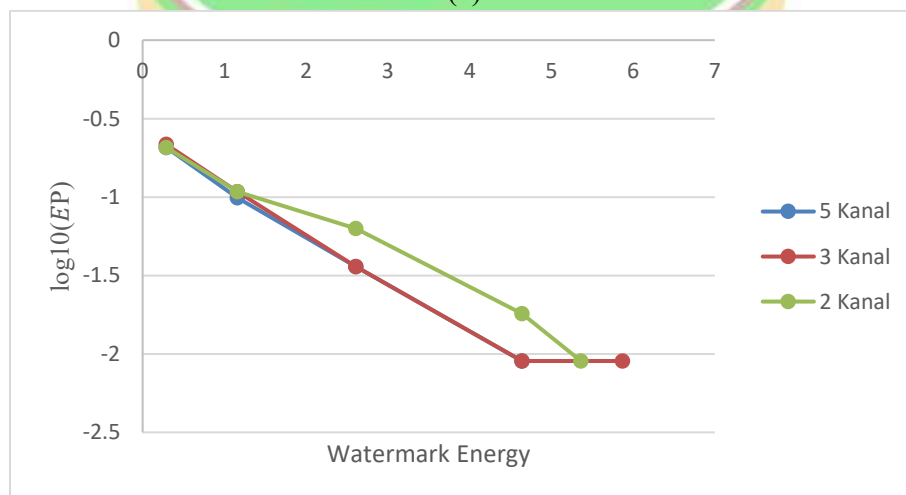
(b)



(c)



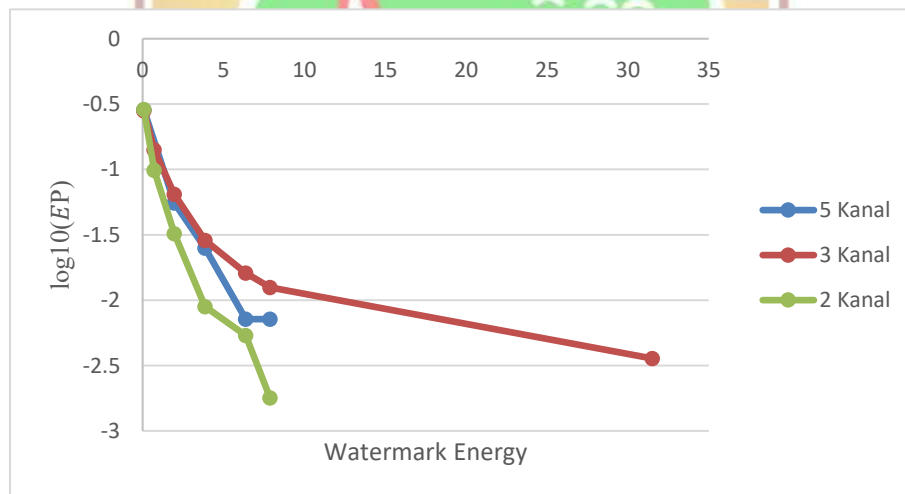
(d)



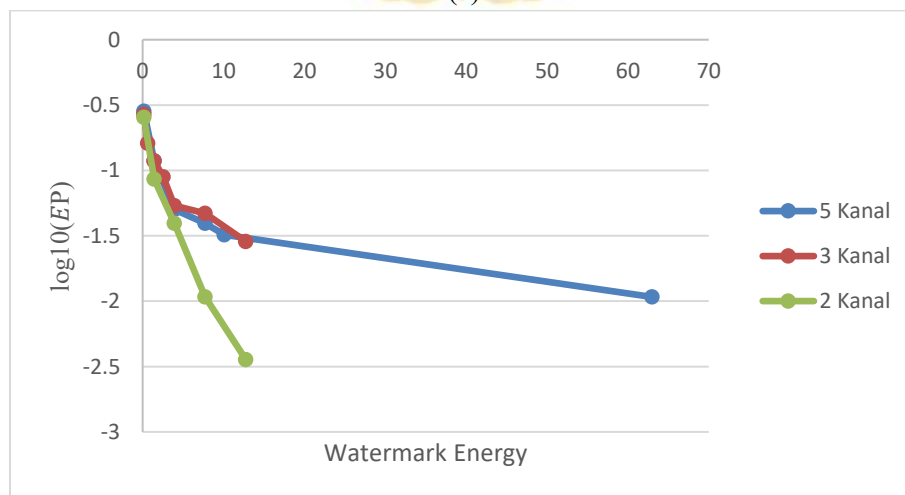
(e)

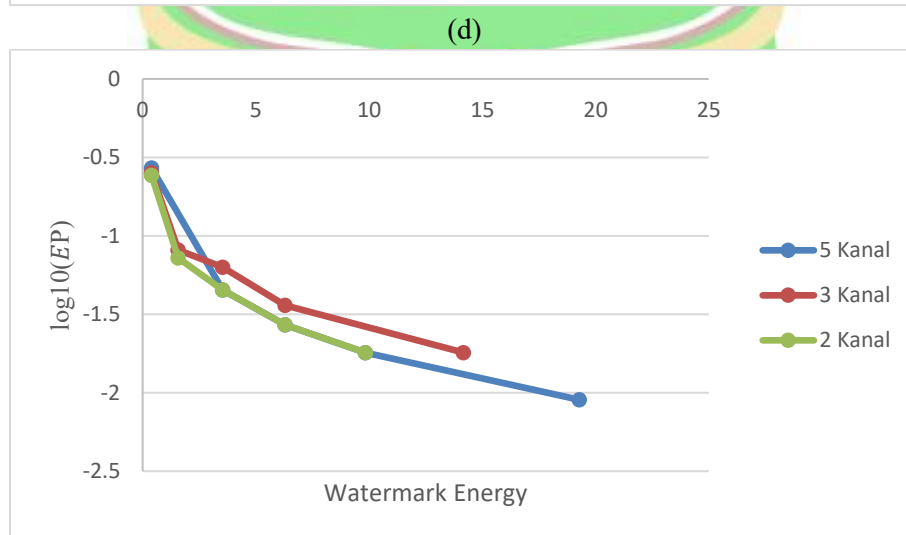
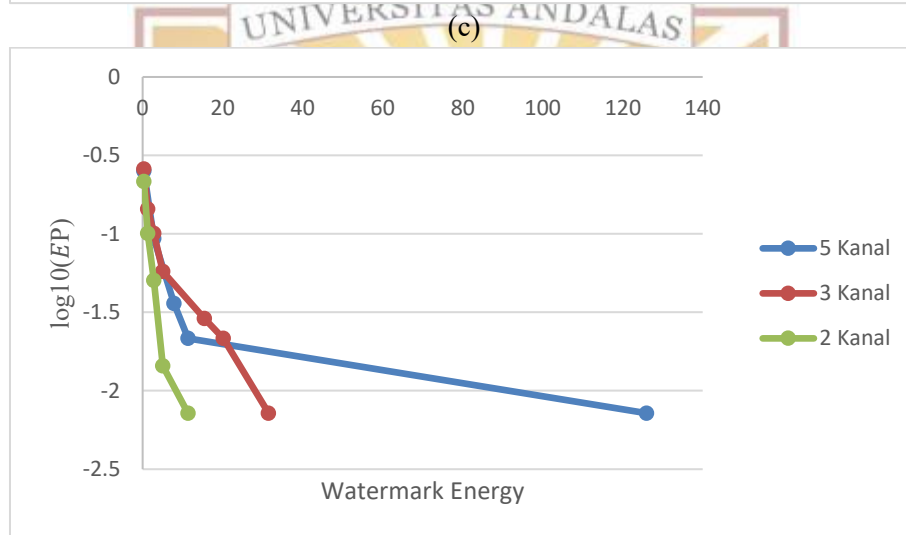
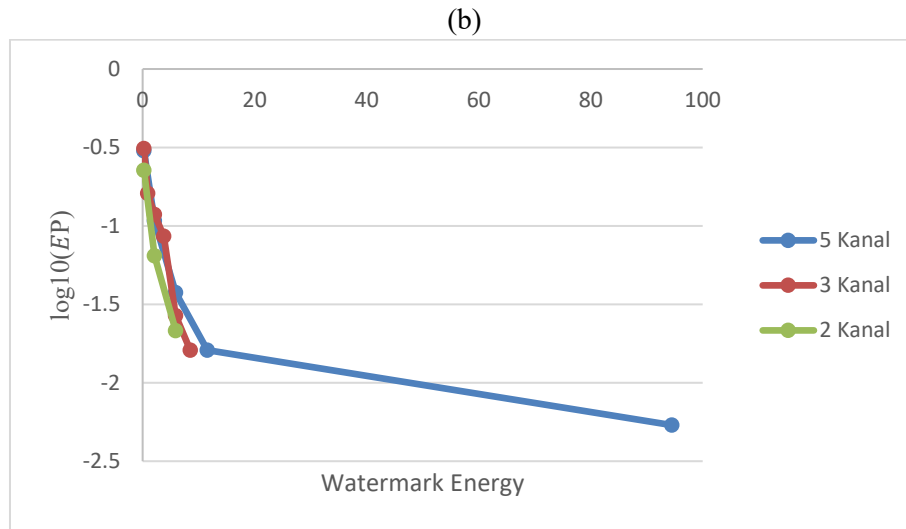
Gambar 4. 1 Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sebelum *downmix* sampel audio *Applause* dengan (a) $N = 1024$, (b) $N = 2048$, (c) $N = 3072$, (d) $N = 4096$, dan (e) $N = 5120$

Hasil pengujian pada sampel audio *Applause* menunjukkan fenomena titik balik performa yang sangat bergantung pada ukuran *frame*. Pada Gambar 4.1(a) dan (b), dengan ukuran *frame* pendek ($N=1024$ dan $N=2048$), mekanisme 2 kanal dengan 1 level hierarki terbukti paling superior karena mampu mencapai nilai $\log_{10}(EP)$ sebesar -2.75 pada WE 5.7 , sementara mekanisme 5 kanal masih membutuhkan WE 4.6 hanya untuk menyentuh nilai $\log_{10}(EP)$ sebesar -1.8 . Keunggulan hierarki rendah ini mulai memudar saat memasuki $N=3072$ (Gambar 4.1(c)), di mana mekanisme 3 kanal mulai lebih efisien dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.25 pada WE 5.2 dibandingkan mekanisme 2 kanal yang membutuhkan WE 6.2 untuk tingkat eror yang sama. Pada Gambar 4.1(d) dan (e), pergeseran performa menjadi sangat drastis di ukuran *frame* besar ($N=4096$ dan $N=5120$), di mana mekanisme 5 kanal mengambil alih posisi optimal dengan capaian nilai $\log_{10}(EP)$ sebesar -2.1 pada WE 5.6 , sedangkan mekanisme 2 kanal justru mengalami saturasi dan tertahan di $\log_{10}(EP)$ sebesar -1.8 meskipun energi dinaikkan hingga 5.7 . Kondisi ini mencapai puncaknya pada $N=5120$ di mana mekanisme 5 kanal dan 3 kanal bergerak beriringan mencapai nilai $\log_{10}(EP)$ sebesar -2.05 pada kisaran WE 4.6 hingga 5.8 .



(a)

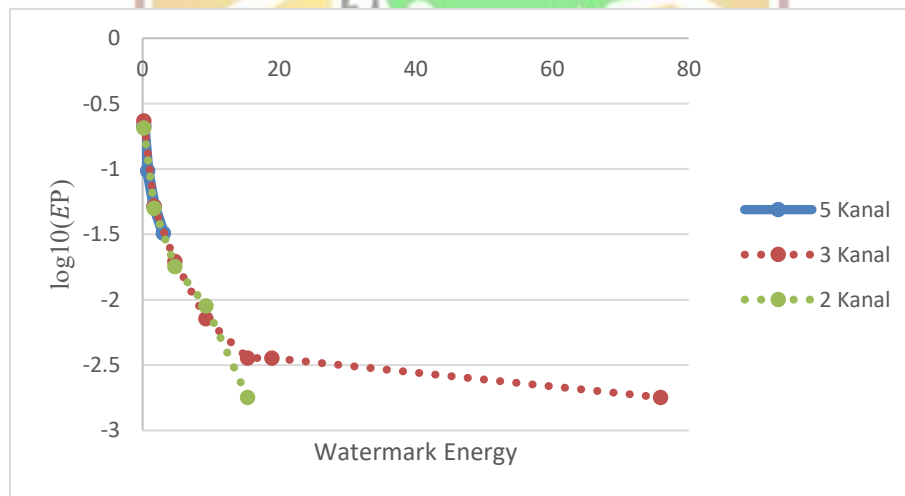




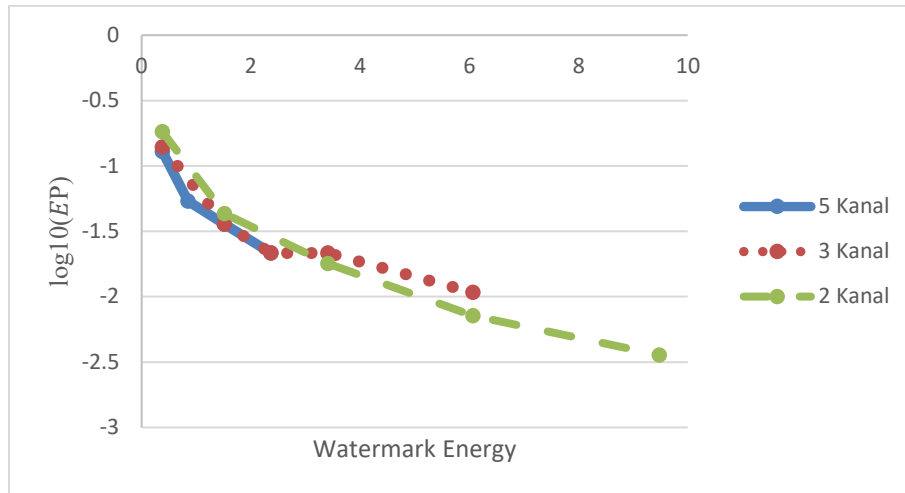
(e)

Gambar 4.2 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sebelum *downmix* sampel audio *Vivaldi* dengan (a) N = 1024, (b) N = 2048, (c) N = 3072, (d) N = 4096, dan (e) N = 5120

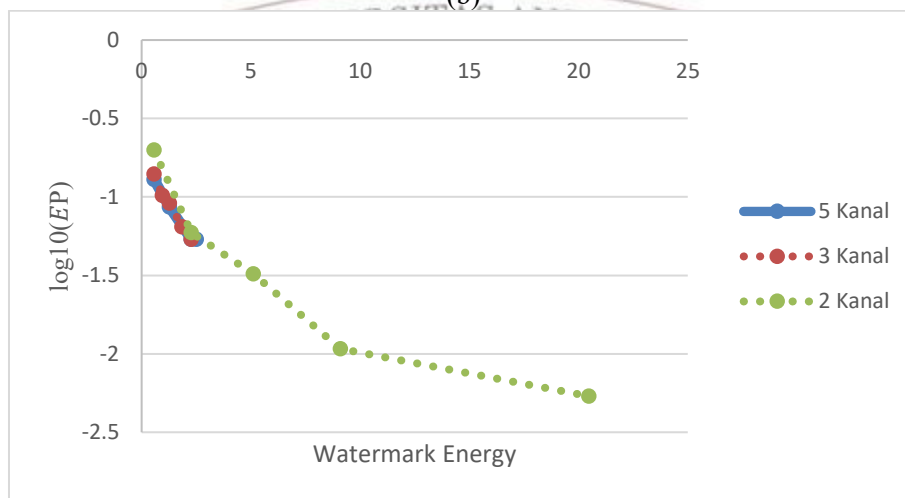
Hasil pengujian pada sampel *Vivaldi* menunjukkan dominasi mutlak mekanisme 2 kanal (1 level hierarki) di seluruh variasi ukuran *frame*. Pada $N=1024$ (Gambar 4.2(a)), mekanisme ini sangat efisien dengan mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.0 pada WE 3.8 dan terus menurun hingga $\log_{10}(\text{EP})$ sebesar -2.75 pada WE 7.8, jauh mengungguli mekanisme 5 kanal yang membutuhkan WE 7.0 hanya untuk mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.1. Superioritas ini semakin kontras pada $N=2048$ (Gambar 4.2(b)), di mana mekanisme 2 kanal mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.5 pada WE 12.5, sementara mekanisme 5 kanal menunjukkan inefisiensi ekstrem dengan menuntut WE hingga 62.0 untuk mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.0. Tren inefisiensi hierarki tinggi mencapai puncaknya pada $N=3072$ dan $N=4096$ (Gambar 4.2(c) dan (d)), di mana mekanisme 5 kanal membutuhkan WE masif hingga 125.0 hanya untuk menyamai capaian nilai $\log_{10}(\text{EP})$ sebesar -2.1 milik mekanisme 2 kanal yang hanya memerlukan WE 9.0. Meskipun pada $N=5120$ (Gambar 4.2(e)) jarak performa sedikit menyempit dengan mekanisme 5 kanal yang mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.0 pada WE 19.0, mekanisme 2 kanal tetap menjadi pilihan paling konsisten dengan capaian nilai $\log_{10}(\text{EP})$ sebesar -1.75 pada WE 10.0. Secara keseluruhan, proses *downmix* bertingkat pada sampel *Vivaldi* bersifat destruktif, sehingga penyederhanaan hierarki menjadi solusi yang efektif untuk menjaga integritas pesan dengan penggunaan energi yang minimal.



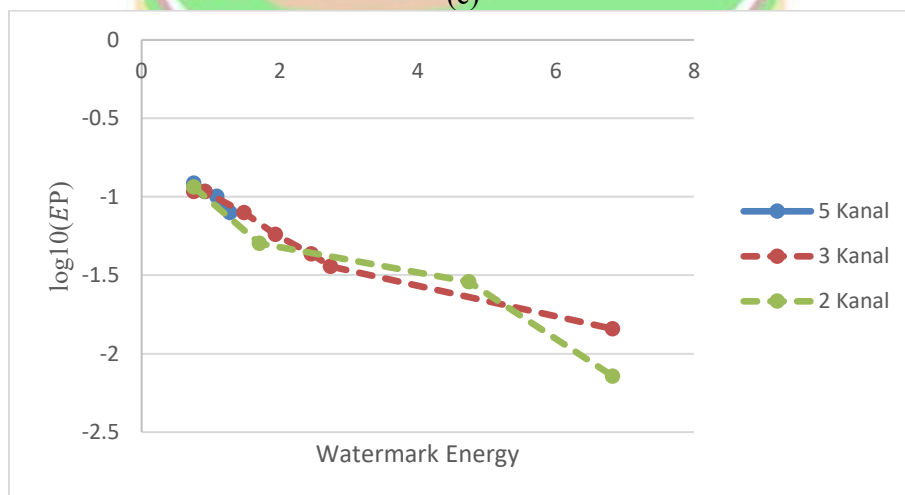
(a)



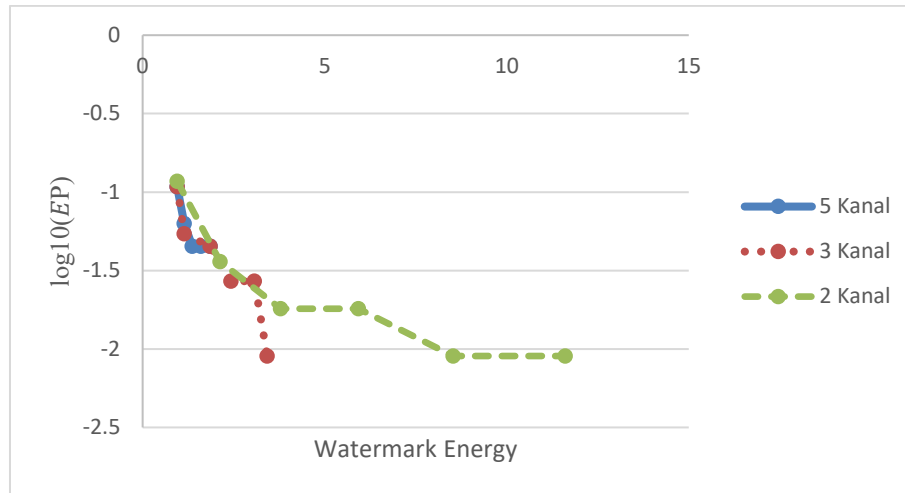
(b)



(c)



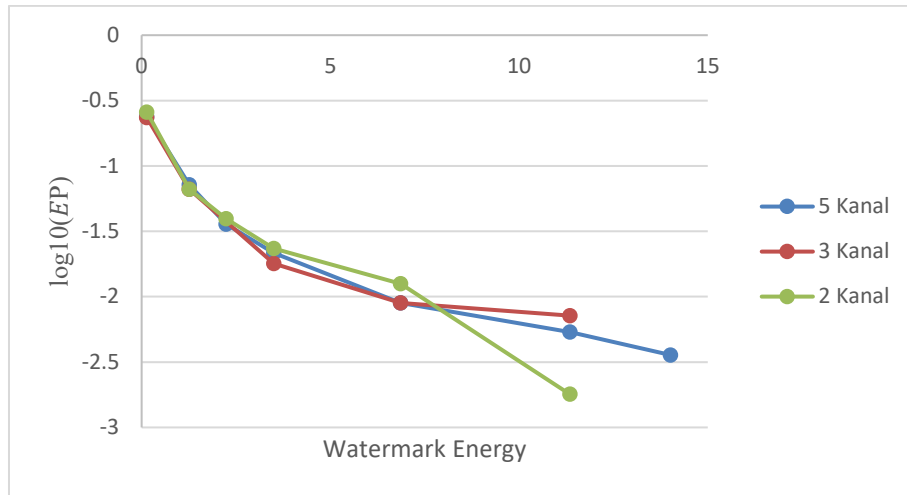
(d)



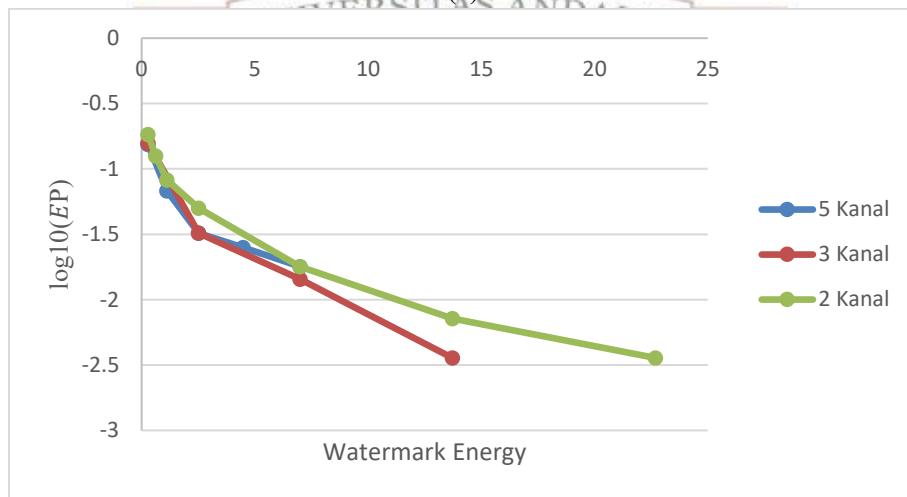
(e)

Gambar 4.3 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sebelum *downmix* sampel audio *Talking* dengan (a) $N = 1024$, (b) $N = 2048$, (c) $N = 3072$, (d) $N = 4096$, dan (e) $N = 5120$

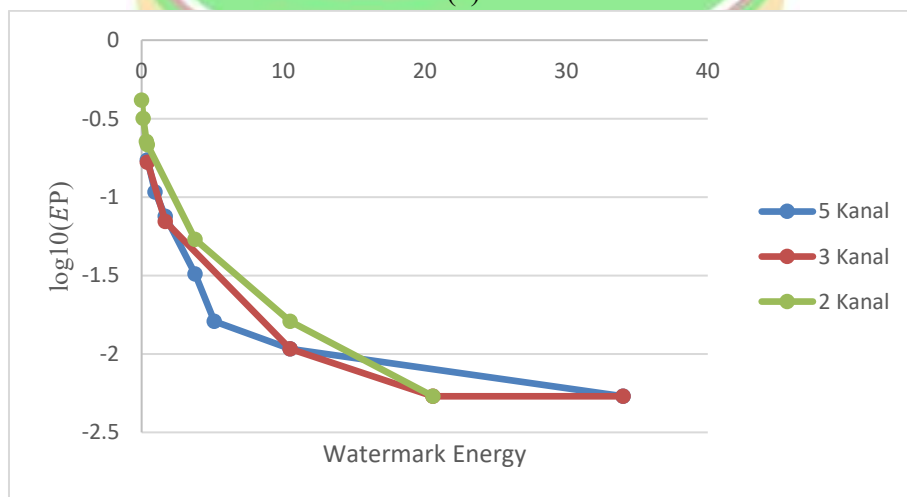
Hasil pengujian pada sampel *Talking* menunjukkan dominasi keunggulan mutlak dari penyederhanaan hierarki melalui mekanisme 2 kanal (1 level). Pada $N=1024$ (Gambar 4.3(a)), mekanisme 2 kanal sangat efisien dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.75 pada WE 15.0 , sementara mekanisme 3 kanal menunjukkan inefisiensi ekstrem karena membutuhkan lonjakan energi hingga 75.0 untuk mencapai tingkat eror yang sama. Memasuki $N=2048$ (Gambar 4.3(b)), terjadi pemulihan efisiensi pada mekanisme 3 kanal yang mampu mencapai nilai $\log_{10}(EP)$ sebesar -2.0 pada WE 6.0 , namun mekanisme 2 kanal tetap lebih unggul dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.45 pada WE hanya 9.5 . Tren ini berlanjut pada $N=3072$ dan $N=4096$ (Gambar 4.3(c) dan (d)), di mana mekanisme 2 kanal menjadi satu-satunya konfigurasi yang konsisten menekan *error* hingga di bawah nilai $\log_{10}(EP)$ sebesar -2.0 , sedangkan mekanisme lainnya sering mengalami saturasi dini di kisaran nilai $\log_{10}(EP)$ sebesar -1.3 . Pada ukuran *frame* terbesar $N=5120$ (Gambar 4.3(e)), meski mekanisme 3 kanal sempat turun cepat di fase awal, mekanisme 2 kanal tetap menjadi opsi paling stabil dengan capaian akhir nilai $\log_{10}(EP)$ sebesar -2.05 pada WE 11.5 . Secara keseluruhan, sampel *Talking* mengukuhkan bahwa struktur hierarki sederhana jauh lebih andal dan hemat energi dalam mencapai target akurasi tinggi dibandingkan struktur kompleks yang rentan terhadap inefisiensi energi.



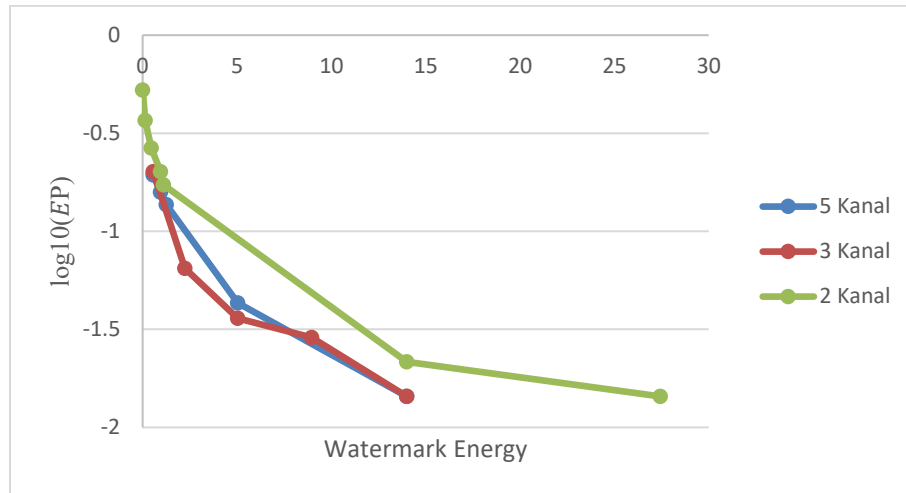
(a)



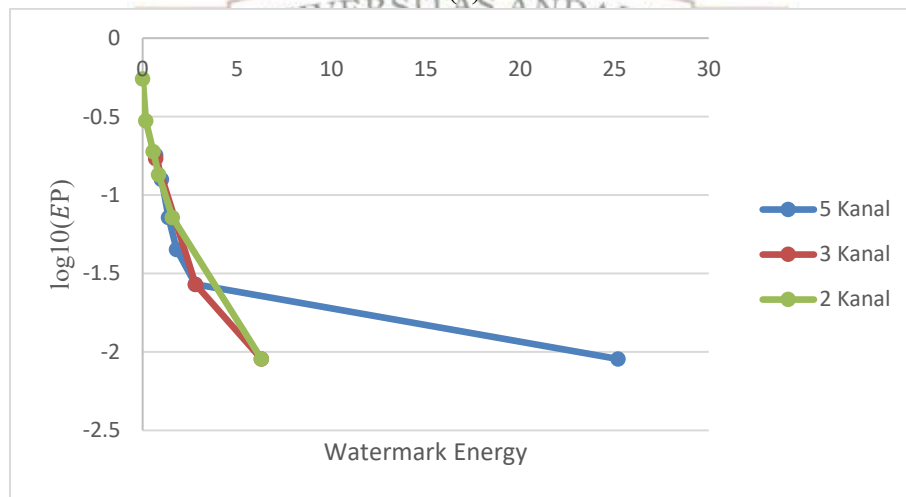
(b)



(c)



(d)



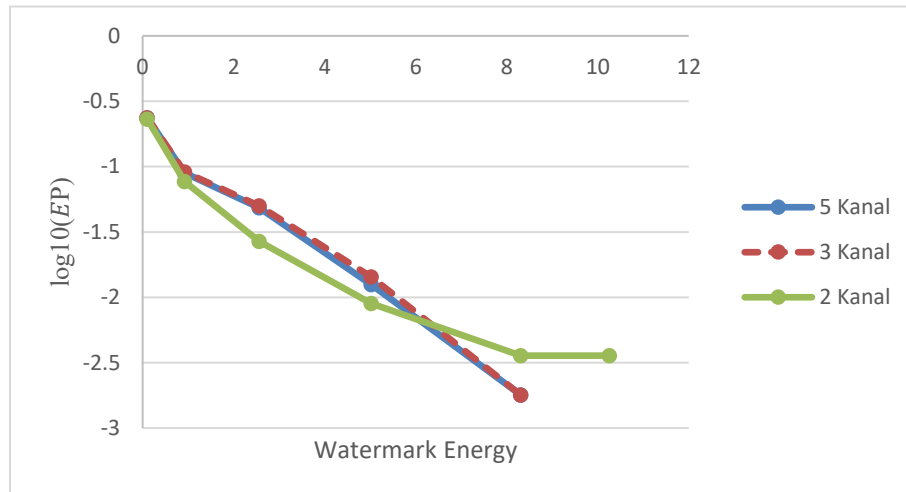
(e)

Gambar 4.4 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sebelum *downmix* sampel audio *Drum* dengan (a) N = 1024, (b) N = 2048, (c) N = 3072, (d) N = 4096, dan (e) N = 5120

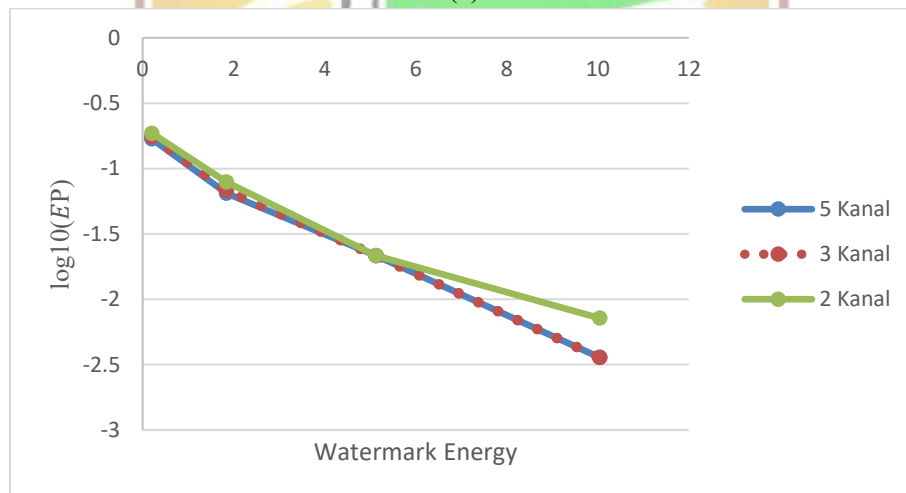
Hasil pengujian pada sampel *Drum* menunjukkan fluktuasi efisiensi yang menarik, namun tetap didominasi oleh mekanisme penyederhanaan hierarki. Pada N=1024 (Gambar 4.4(a)), mekanisme 2 kanal (1 level) paling efisien dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.75 pada WE 11.5, mengungguli mekanisme 3 kanal yang hanya mencapai nilai $\log_{10}(EP)$ sebesar -2.1 pada energi yang sama. Meskipun pada N=2048 dan N=4096 (Gambar 4.4(b) dan (d)) mekanisme 3 kanal sempat mengambil alih keunggulan di rentang energi tinggi (mencapai nilai $\log_{10}(EP)$ sebesar -1.85 pada WE 14.0), mekanisme 2 kanal kembali menunjukkan dominasi telak pada ukuran *frame* menengah dan besar.

Pada N=3072 (Gambar 4.4(c)), mekanisme 2 kanal mencapai nilai $\log_{10}(EP)$ sebesar -2.25 hanya dengan WE 20.5, jauh lebih hemat dibandingkan mekanisme lainnya yang membutuhkan WE hingga 34.0. Superioritas ini semakin

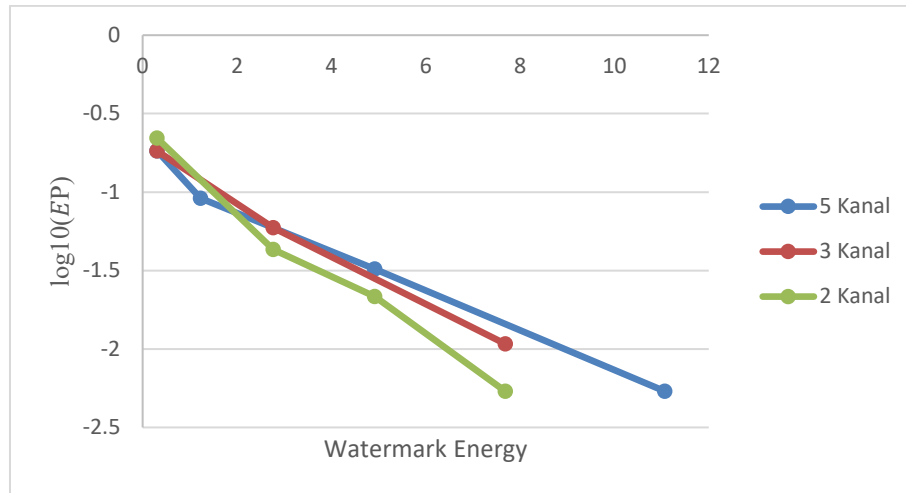
nyata pada ukuran *frame* terbesar $N=5120$ (Gambar 4.4(e)), di mana mekanisme 2 kanal mampu menyentuh nilai $\log_{10}(EP)$ sebesar -2.0 hanya dengan WE 6.5, sedangkan mekanisme 5 kanal memerlukan energi masif hingga 25.0 untuk hasil yang sama. Secara keseluruhan, sampel audio *Drum* mengonfirmasi bahwa simplifikasi hierarki menjadi 2 kanal merupakan strategi paling konsisten dan hemat energi dalam mempertahankan integritas pesan di berbagai variasi ukuran *frame*.



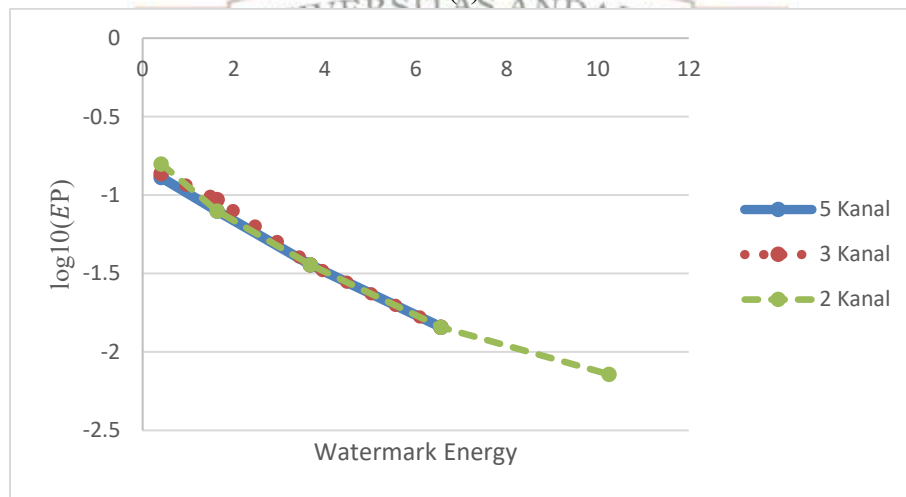
(a)



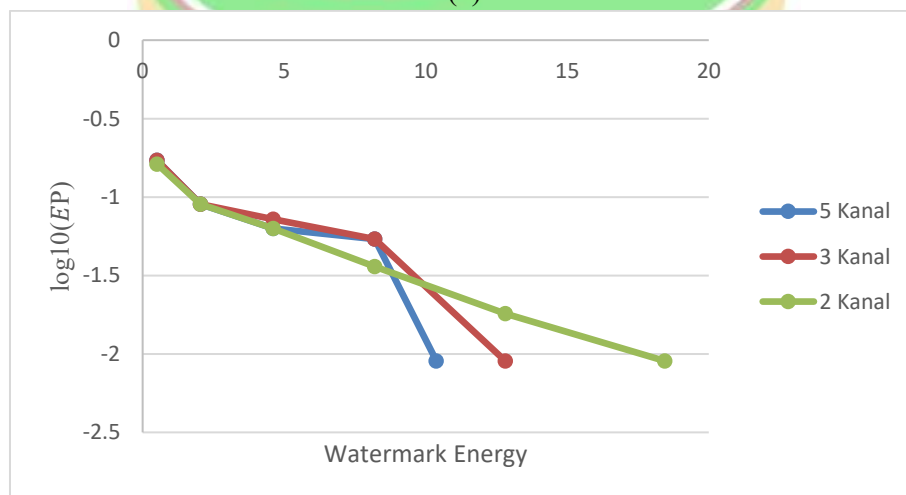
(b)



(c)



(d)



(e)

Gambar 4.5 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sebelum *downmix* sampel audio *Laughter* dengan (a) $N = 1024$, (b) $N = 2048$, (c) $N = 3072$, (d) $N = 4096$, dan (e) $N = 5120$

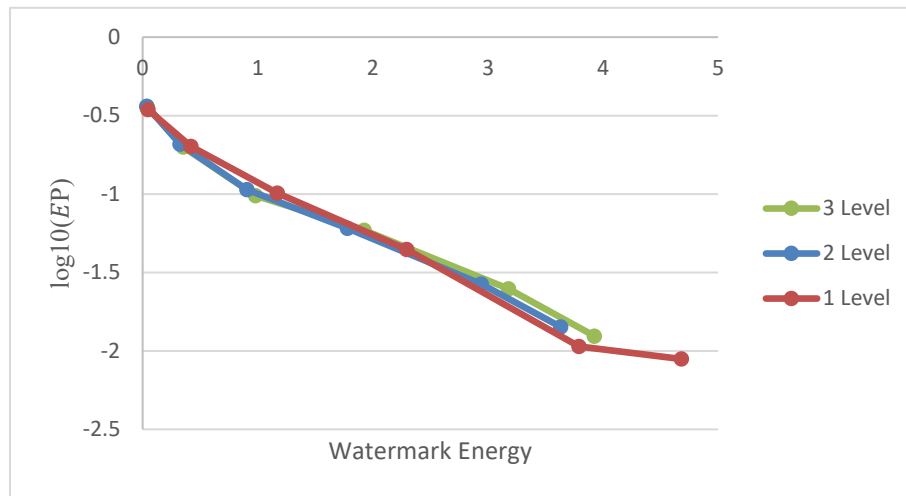
Hasil pengujian pada sampel *Laughter* menunjukkan pergeseran efisiensi yang dinamis seiring perubahan ukuran *frame*. Pada ukuran *frame* pendek $N=1024$ dan $N=2048$ (Gambar 4.5(a) dan (b)), mekanisme 3 kanal tampil paling efisien dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.75 pada WE 8.3 ($N=1024$) dan -2.45 pada WE 10.0 ($N=2048$), mengungguli mekanisme 2 kanal yang membutuhkan energi lebih besar untuk tingkat *error* yang sama. Dominasi beralih ke mekanisme 2 kanal pada $N=3072$ dan $N=4096$. Pada $N=3072$ (Gambar 4.5(c)), mekanisme 2 kanal mencapai nilai $\log_{10}(EP)$ sebesar -2.25 hanya dengan WE 7.6, sementara mekanisme 5 kanal membutuhkan WE 11.0 untuk hasil serupa. Puncaknya pada $N=4096$ (Gambar 4.5(d)), mekanisme 2 kanal menjadi satu-satunya konfigurasi yang mampu menembus ambang batas *error* rendah secara signifikan. Namun, pada ukuran *frame* terbesar $N=5120$ (Gambar 4.5(e)), terjadi titik balik di mana mekanisme 5 kanal menemukan momentum terbaiknya dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.0 menggunakan WE hanya ~ 10 , jauh lebih hemat dibandingkan mekanisme 2 kanal yang membutuhkan WE 18.45. Secara keseluruhan, sampel *Laughter* membuktikan bahwa penyederhanaan hierarki efektif pada *frame* menengah, namun distribusi *watermark* ke banyak kanal (5 kanal) menjadi strategi paling optimal saat menggunakan *frame* yang sangat panjang.

Pada kelima sampel audio, tidak terlihat pola penurunan *Error Probability* yang konsisten seiring dengan bertambahnya ukuran *frame*. Salah satu penyebab utama dari ketidak-konsistenan ini disebabkan oleh penggunaan variasi ukuran *frame* yang sama pada seluruh mekanisme, tidak memberikan perbandingan yang setara dari segi beban sinyalnya. Pada mekanisme 5 kanal, ukuran *frame* yang disisipkan mengakumulasi *noise* dari 5 sumber sekaligus. Akibatnya, isi *frame* menjadi sangat padat, sehingga memperpanjang ukuran *frame* tidak menghasilkan nilai yang selalu progresif karena tertutup oleh tumpukan *noise* yang terlalu masif. Hal ini membuat grafik *Error Probability* sulit turun meskipun energi dan ukuran *frame* sudah ditambah. Penyederhanaan level hierarki tidak menghilangkan karakteristik destruktif dari proses penggabungan sinyal itu sendiri. Inkonsistensi tetap terjadi karena interferensi antar-kanal dan interferensi dari audio asli tidak selalu berbanding lurus dengan penambahan energi maupun penyederhanaan hierarki.

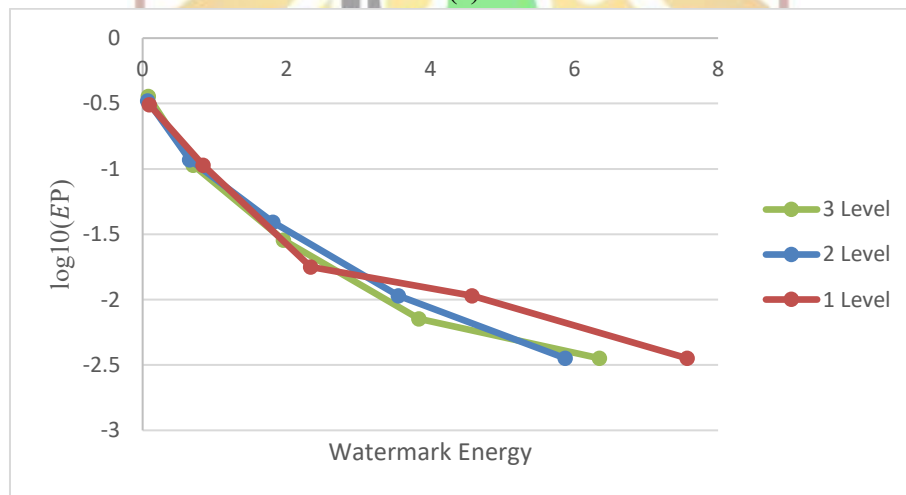
4.2.2. Penyisipan sesudah *Downmix*

Untuk membandingkan performa penyisipan sebelum *downmix*, dilakukan pengujian penyisipan setelah *downmix* dengan konfigurasi yang dapat dilihat pada Gambar 3.2 (d) s.d. 3.2 (f). Pada skenario pengujian ini, audio terlebih dahulu melalui proses *downmixing MPEG Surround* hingga tuntas, sehingga proses penyisipan *watermark* dilakukan hanya pada 1 kanal audio hasil akhir (mono). Perbedaan mekanisme terletak pada kompleksitas rantai penurunan kanal (*downmix*) sebelum penyisipan dilakukan. Pada mekanisme *downmix* 3 level, sinyal

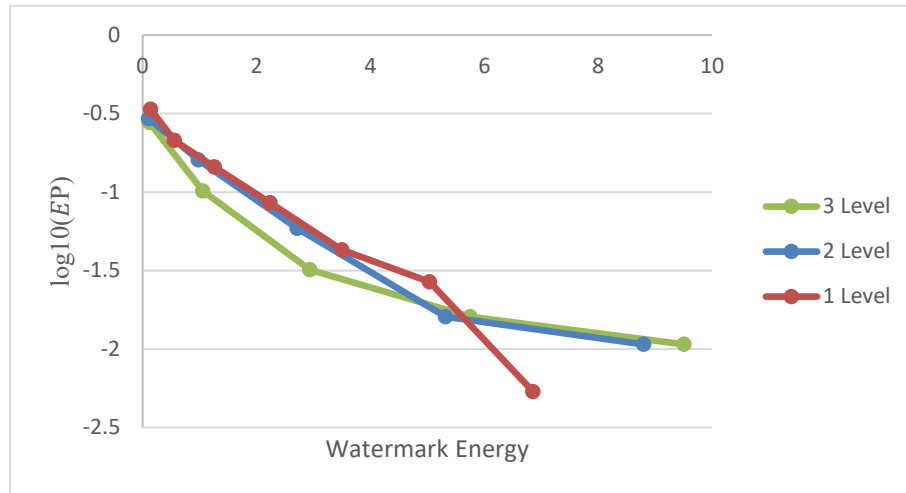
input awal 5 kanal diproses melewati rantai penuh tiga modul R-OTT secara bertingkat untuk diringkas menjadi sinyal mono. Mekanisme ini kemudian disederhanakan menjadi *downmix* 2 level, di mana input diperlakukan sebagai format 3 kanal yang hanya perlu melewati dua tahapan reduksi modul R-OTT untuk mencapai format mono. Penyederhanaan paling akhir diterapkan pada *downmix* 1 level, yang dimulai dari 2 kanal dan hanya melewati satu tahap modul R-OTT tunggal untuk menghasilkan output mono. *Spatial Parameter* yang dihasilkan pada setiap tahap R-OTT tetap diekstraksi dan disimpan untuk keperluan proses *decoding*.



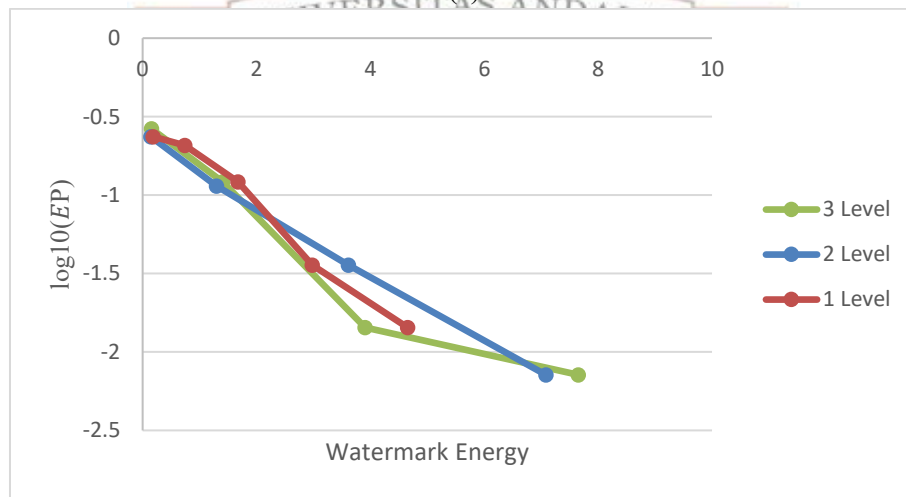
(a)



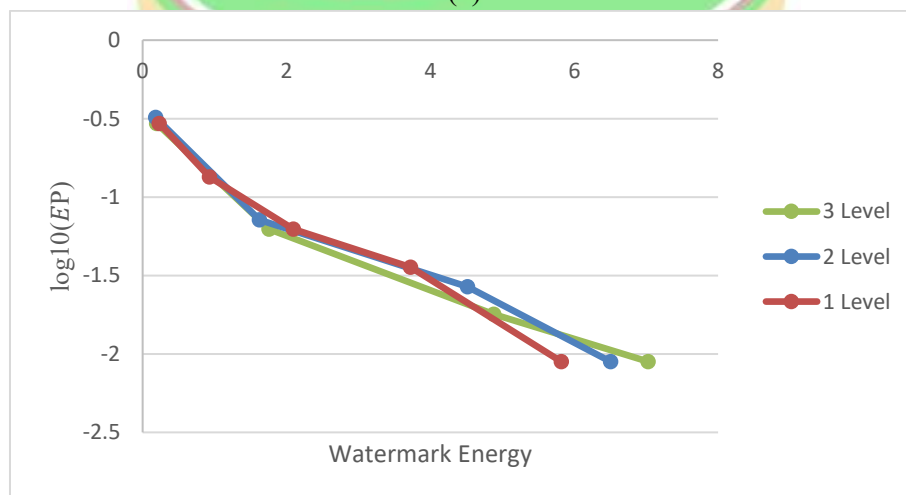
(b)



(c)



(d)

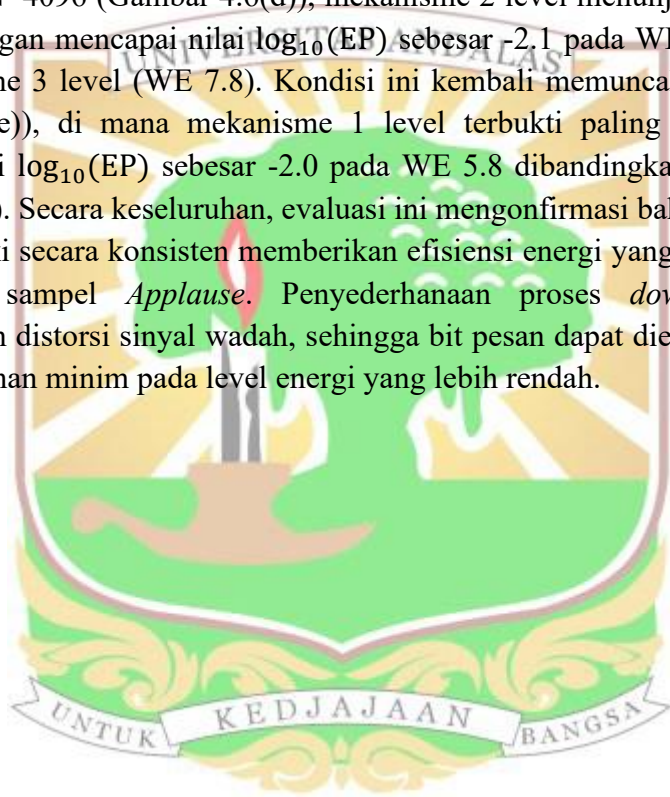


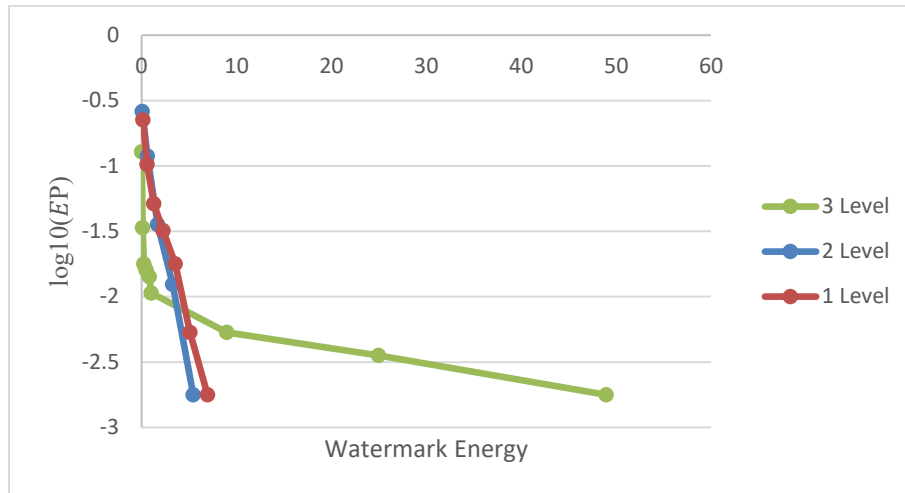
(e)

Gambar 4.6 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sesudah *downmix* sampel audio *Applause* dengan (a) $N = 1024$, (b) $N = 2048$, (c) $N = 3072$, (d) $N = 4096$, dan (e) $N = 5120$

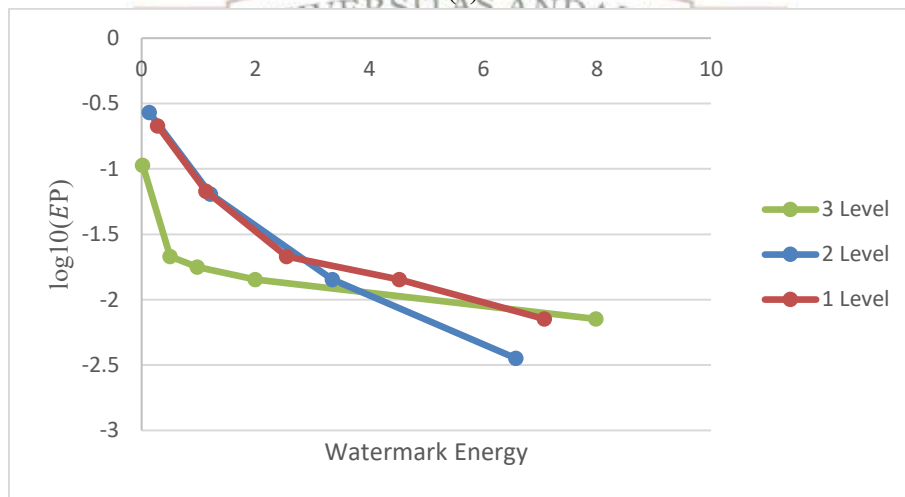
Hasil pengujian pada sampel *Applause* dengan skenario penyisipan sesudah *downmix* menunjukkan bahwa penyederhanaan hierarki cenderung meningkatkan efisiensi ekstraksi. Pada N=1024 (Gambar 4.6(a)), ketiga mekanisme menunjukkan karakteristik identik, di mana mekanisme 1 level sedikit lebih efisien dengan mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.0 pada WE 3.6. Saat ukuran *frame* ditingkatkan menjadi N=2048 (Gambar 4.6(b)), terjadi separasi performa di mana mekanisme 2 level sempat unggul mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.0 pada WE 3.6, namun dominasi kembali beralih ke mekanisme 1 level pada N=3072 (Gambar 4.6(c)) dengan capaian nilai $\log_{10}(\text{EP})$ sebesar -2.25 pada WE 6.8, jauh lebih responsif dibandingkan mekanisme 3 level yang membutuhkan WE 9.7 untuk target nilai $\log_{10}(\text{EP})$ yang lebih rendah (sebesar -2.0).

Pada N=4096 (Gambar 4.6(d)), mekanisme 2 level menunjukkan stabilitas yang baik dengan mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.1 pada WE 7.1, sedikit di atas mekanisme 3 level (WE 7.8). Kondisi ini kembali memuncak pada N=5120 (Gambar 4.6(e)), di mana mekanisme 1 level terbukti paling efisien dengan mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.0 pada WE 5.8 dibandingkan mekanisme 2 level (WE 6.3). Secara keseluruhan, evaluasi ini mengonfirmasi bahwa konfigurasi 1 level hierarki secara konsisten memberikan efisiensi energi yang lebih baik atau setara pada sampel *Applause*. Penyederhanaan proses *downmix* terbukti meminimalkan distorsi sinyal wadah, sehingga bit pesan dapat diekstraksi dengan tingkat kesalahan minim pada level energi yang lebih rendah.

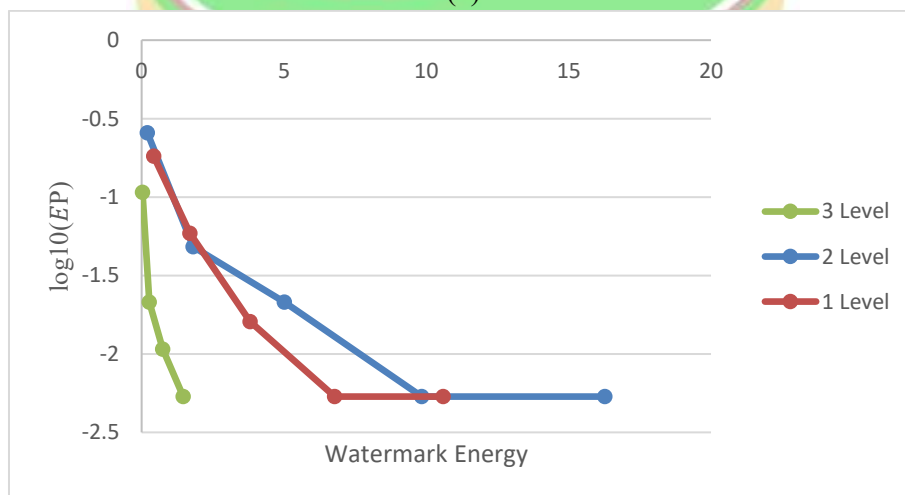




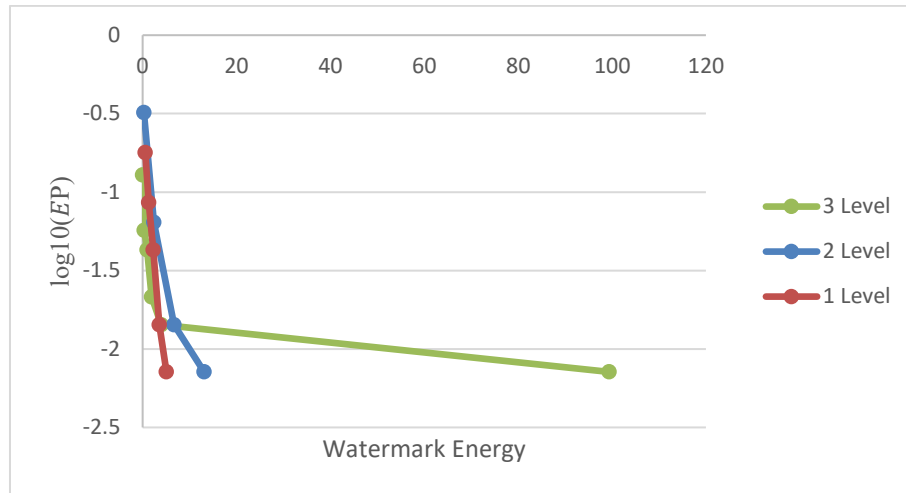
(a)



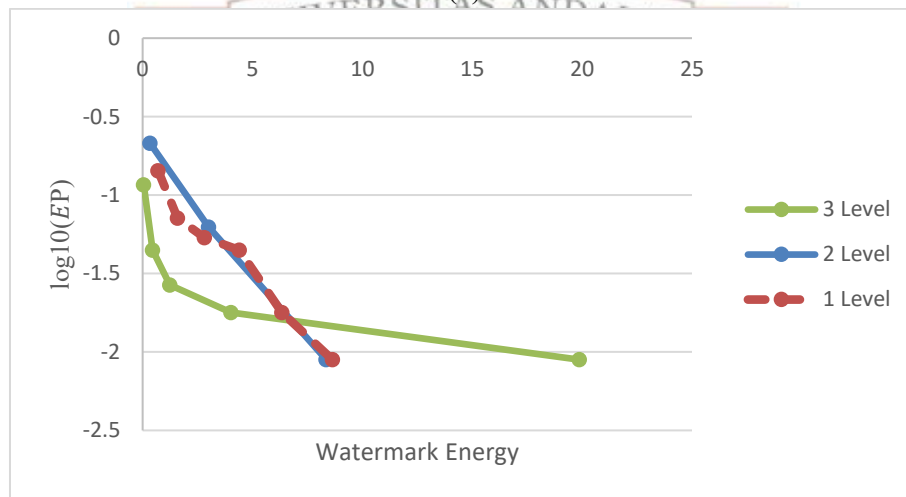
(b)



(c)



(d)

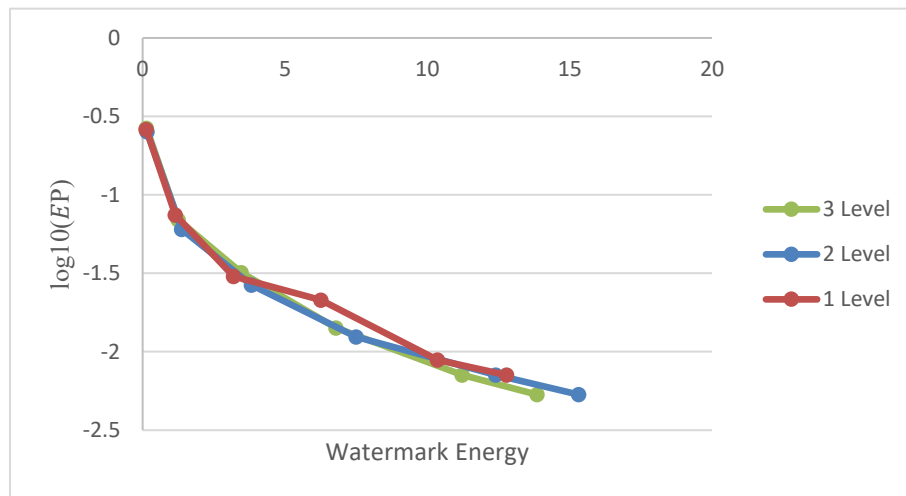


(e)

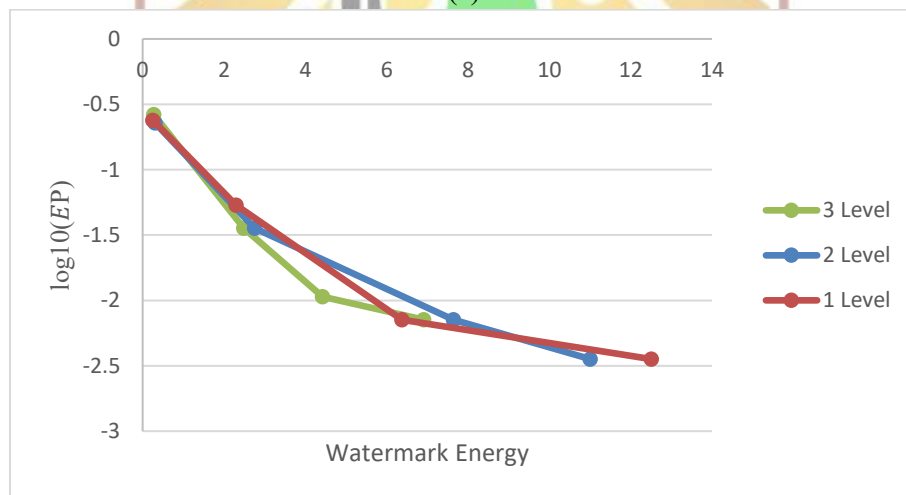
Gambar 4.7 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sesudah *downmix* sampel audio *Vivaldi* dengan (a) $N = 1024$, (b) $N = 2048$, (c) $N = 3072$, (d) $N = 4096$, dan (e) $N = 5120$

Hasil pengujian pada sampel *Vivaldi* dengan skenario penyisipan sesudah *downmix* menunjukkan adanya karakteristik performa yang bergantung pada ukuran *frame*. Pada *frame* pendek $N=1024$ dan $N=2048$ (Gambar 4.7(a) dan (b)), mekanisme 2 level hierarki lebih efisien, di mana pada $N=1024$ mekanisme ini mencapai nilai $\log_{10}(EP)$ sebesar -2.75 dengan WE 5.5 , mengungguli mekanisme 1 level (WE 7.0) dan 3 level (WE 48.9). Namun, terjadi titik balik pada ukuran *frame* menengah ($N=3072$ dan $N=4096$). Pada $N=3072$ (Gambar 4.7(c)), mekanisme 1 level tampil superior dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.3 pada WE hanya 6.0 , sementara mekanisme 2 level membutuhkan energi hingga 16.0 untuk hasil yang sama. Dominasi ini semakin stabil pada $N=4096$ (Gambar 4.7(d)), di mana mekanisme 1 level mencapai nilai $\log_{10}(EP)$ sebesar -2.2 dengan WE hemat sebesar 5.0 , berbanding terbalik dengan mekanisme 3 level yang menuntut energi mendekati 100 .

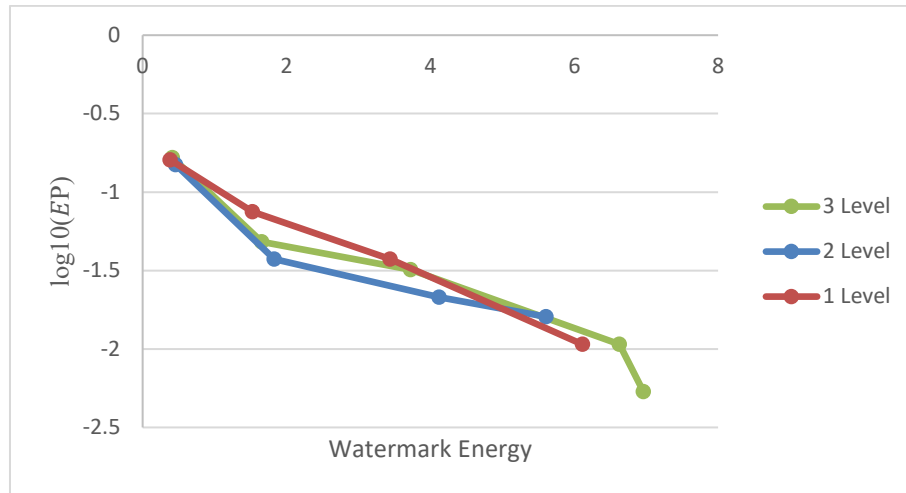
Pada ukuran *frame* terbesar $N=5120$ (Gambar 4.7(e)), performa mekanisme 1 level dan 2 level cenderung konvergen atau bergerak beriringan, dengan kebutuhan energi masing-masing sekitar 8.5 dan 8.2 untuk mencapai nilai $\log_{10}(EP)$ sebesar -2.0. Hal ini menandakan bahwa pada *frame* terpanjang, karakteristik statistik sinyal mono dari kedua jenis *downmix* menjadi serupa. Secara keseluruhan, analisis ini menyimpulkan bahwa penyederhanaan proses *downmix* menjadi 1 level sangat direkomendasikan pada ukuran *frame* menengah hingga besar untuk sampel *Vivaldi* karena mampu menjaga integritas data dengan penggunaan energi yang jauh lebih rendah.



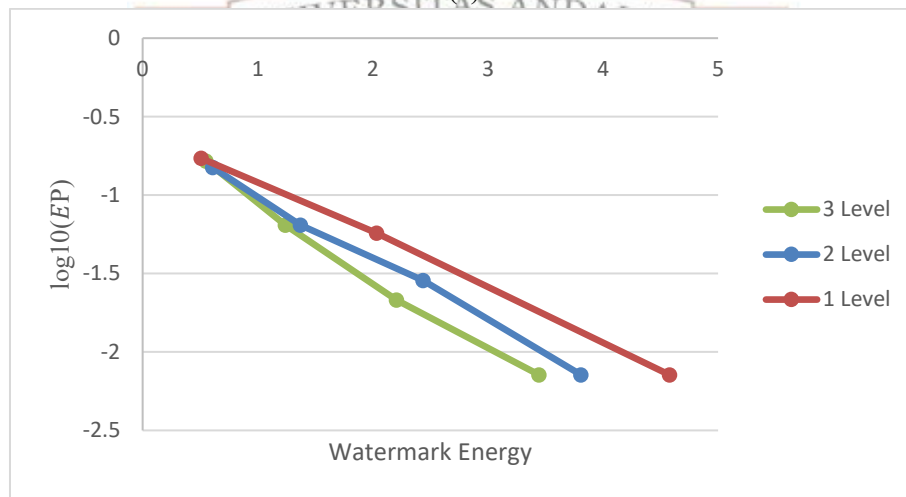
(a)



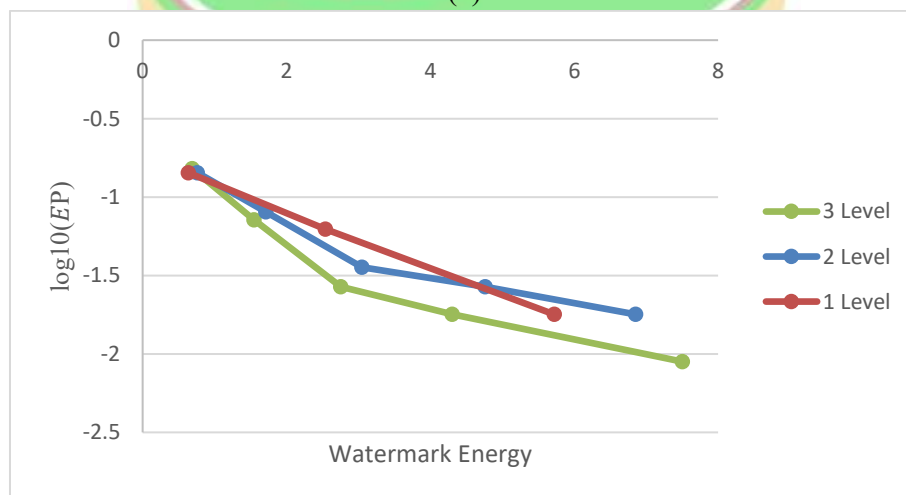
(b)



(c)



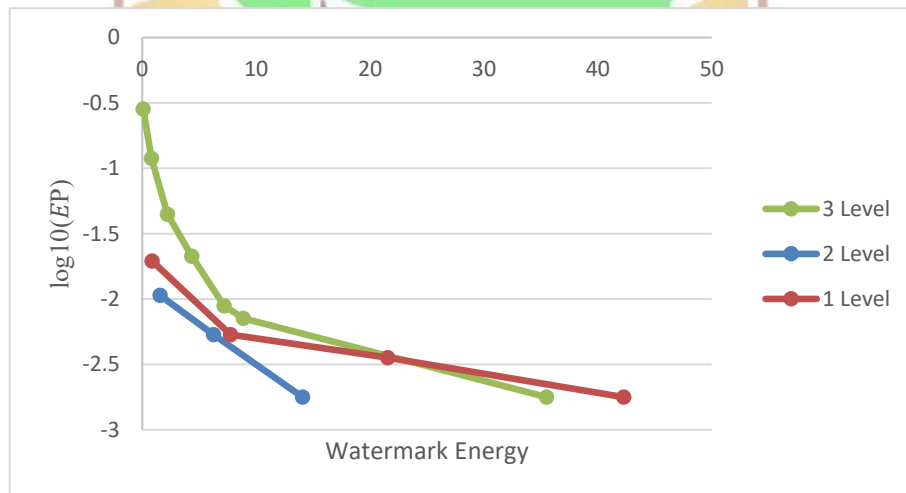
(d)



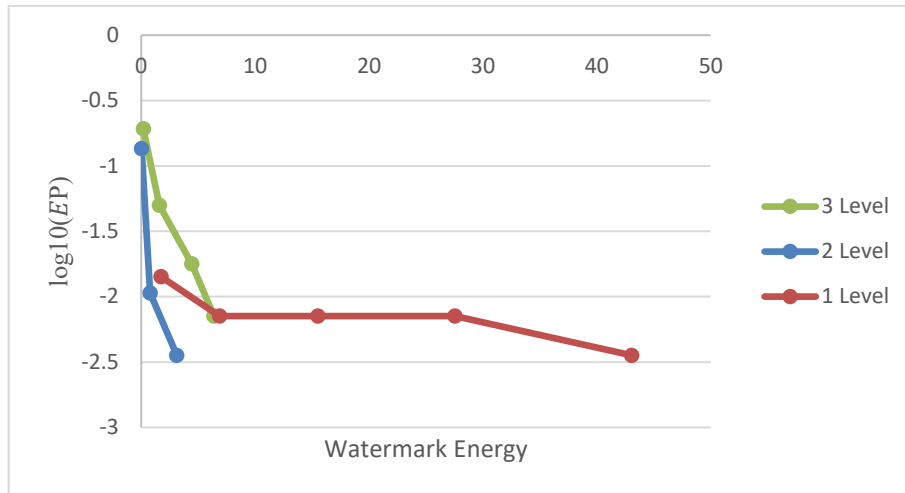
(e)

Gambar 4.8 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sesudah *downmix* sampel audio *Talking* dengan (a) $N = 1024$, (b) $N = 2048$, (c) $N = 3072$, (d) $N = 4096$, dan (e) $N = 5120$

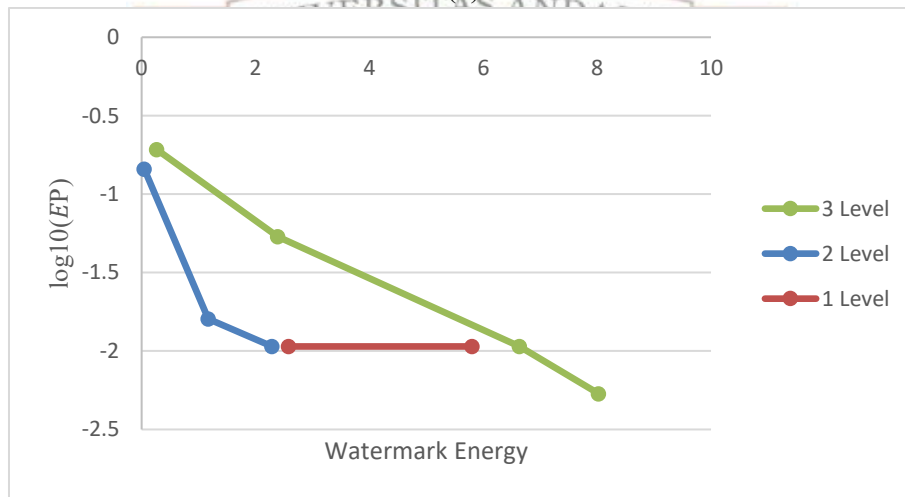
Hasil pengujian pada sampel *Talking* dengan skenario penyisipan sesudah *downmix* menunjukkan karakteristik performa yang dinamis di setiap ukuran *frame*. Pada $N=1024$ (Gambar 4.8(a)), ketiga mekanisme memiliki performa identik dengan nilai $\log_{10}(EP)$ sebesar -1.5 pada WE 3.8. Efisiensi mulai meningkat signifikan pada $N=2048$ (Gambar 4.8(b)), di mana mekanisme 1 level lebih unggul dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.15 pada WE 6.3, dibandingkan mekanisme 2 level yang membutuhkan WE 7.6. Keunggulan mekanisme 1 level berlanjut pada $N=3072$ (Gambar 4.8(c)), di mana mekanisme ini mampu menembus nilai $\log_{10}(EP)$ sebesar -2.0 dengan WE 6.1, sementara mekanisme 2 level mulai mengalami saturasi. Pembalikan tren sempat terjadi pada $N=4096$ (Gambar 4.8(d)), di mana mekanisme 3 level tampil paling agresif dan hemat energi dengan mencapai nilai $\log_{10}(EP)$ sebesar -2.1 pada WE hanya 3.4. Namun, pada ukuran frame terbesar $N=5120$ (Gambar 4.8(e)), mekanisme 1 level kembali memimpin dengan mencapai nilai $\log_{10}(EP)$ sebesar -1.75 pada WE 5.7, mengungguli mekanisme 2 level yang membutuhkan WE 6.9. Secara keseluruhan, penyederhanaan menjadi 1 level hierarki pada sampel *Talking* memberikan performa paling kompetitif dan hemat energi, terutama pada penggunaan frame menengah hingga besar, karena mampu menekan tingkat kesalahan secara lebih konsisten dibandingkan struktur hierarki yang lebih kompleks.



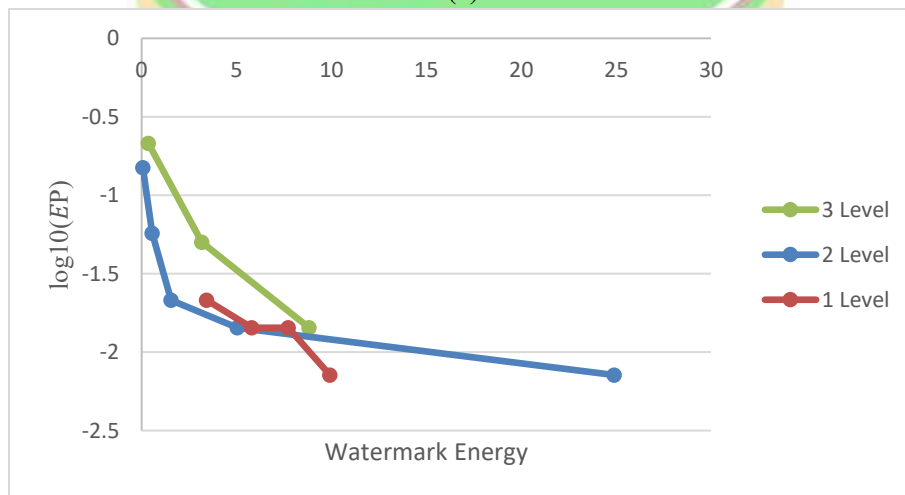
(a)



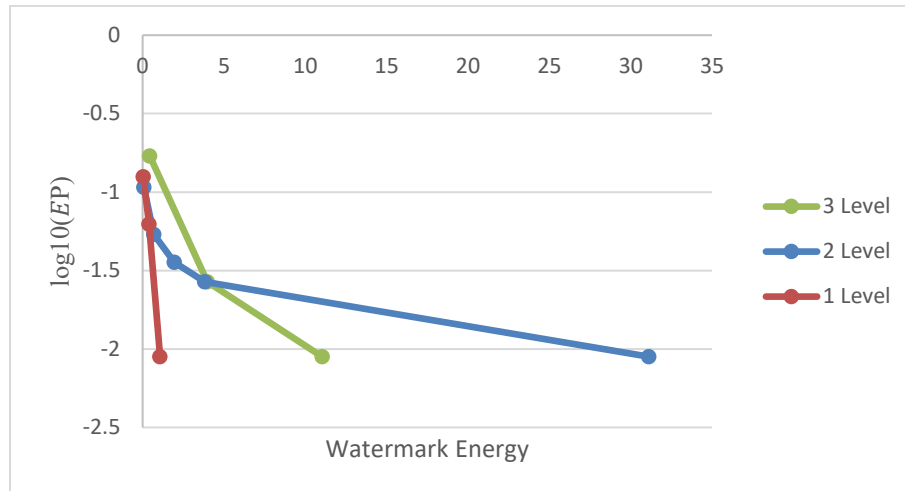
(b)



(c)



(d)

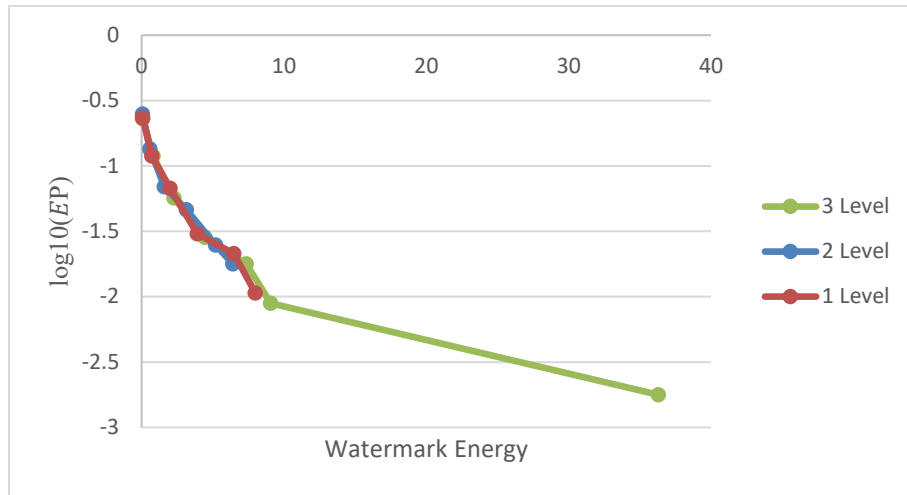


(e)

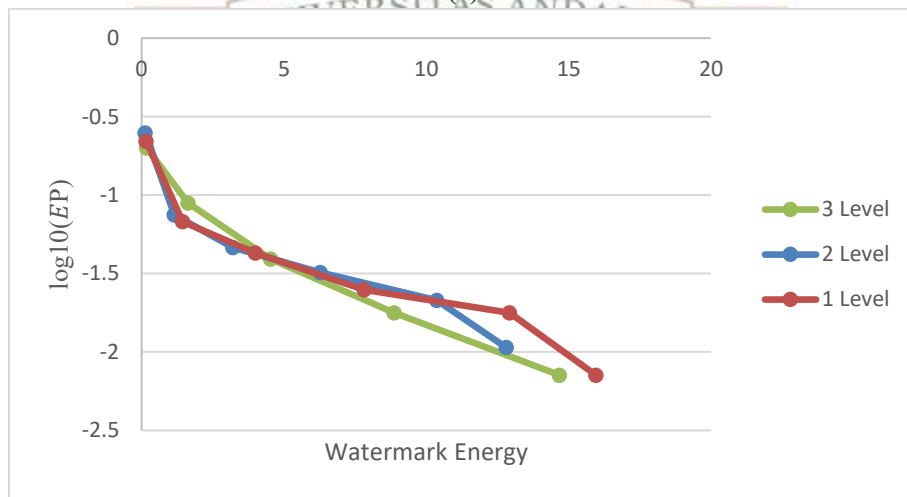
Gambar 4.9 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sesudah *downmix* sampel audio *Drum* dengan (a) N = 1024, (b) N = 2048, (c) N = 3072, (d) N = 4096, dan (e) N = 5120

Hasil pengujian pada sampel *Drum* dengan skenario penyisipan sesudah *downmix* menunjukkan fenomena efisiensi yang bergantung pada ukuran *frame*. Pada *frame* kecil N=1024 dan N=2048 (Gambar 4.9(a) dan (b)), mekanisme 2 level hierarki sangat dominan, pada N=1024 mekanisme ini mencapai nilai log₁₀(EP) sebesar -2.0 dengan WE hanya 1.8, jauh lebih efisien dibanding 1 level (WE 5.0). Dominasi ini menguat pada N=2048 di mana 2 level mencapai nilai log₁₀(EP) sebesar -2.45 pada WE 3.0, sementara 1 level membutuhkan energi ekstrem hingga 43.0.

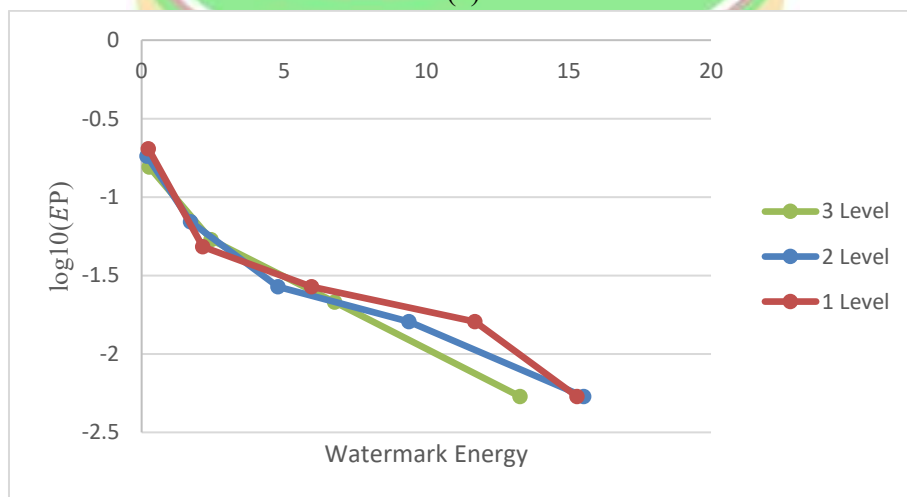
Peralihan tren mulai terlihat pada N=3072 (Gambar 4.9(c)) sebagai titik balik perbaikan performa mekanisme 1 level. Keunggulan ini berbalik sepenuhnya pada *frame* besar N=4096 dan N=5120. Pada N=4096 (Gambar 4.9(d)), mekanisme 1 level mencapai nilai log₁₀(EP) sebesar -2.15 dengan WE 7.6, jauh mengungguli 2 level yang membutuhkan WE 24.8. Tren ini mencapai puncaknya pada N=5120 (Gambar 4.9(e)), di mana mekanisme 1 level menunjukkan superioritas mutlak dengan mencapai nilai log₁₀(EP) sebesar -2.05 pada WE sangat minim (~1.8), sementara mekanisme 2 level menuntut energi hingga 30.8. Secara keseluruhan, analisis ini menyimpulkan bahwa sampel *Drum*, *frame* besar lebih optimal dipadukan dengan *downmix* sederhana (1 level), sedangkan *frame* kecil lebih efektif menggunakan struktur *downmix* 2 level.



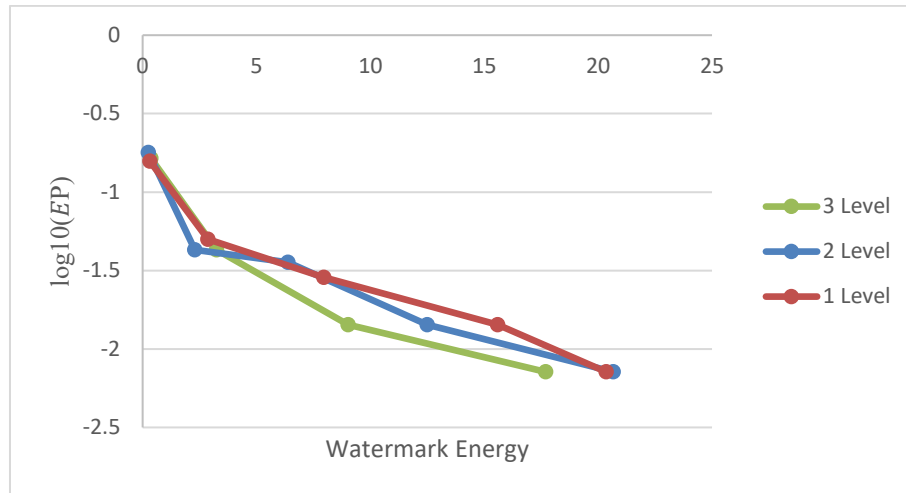
(a)



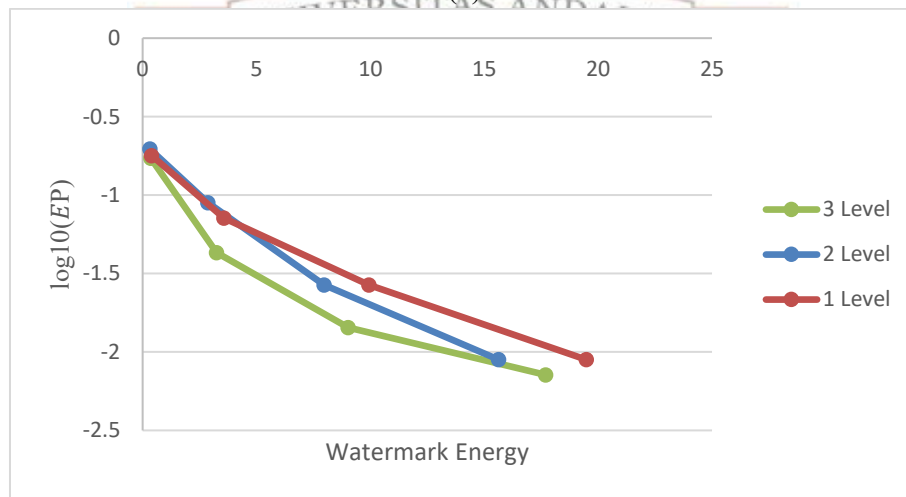
(b)



(c)



(d)



(e)

Gambar 4. 10 (a) s.d. (e) Grafik analisis perbandingan WE dan EP pada setiap level hierarki mekanisme penyisipan sesudah *downmix* sampel audio *Laughter* dengan (a) N = 1024, (b) N = 2048, (c) N = 3072, (d) N = 4096, dan (e) N = 5120

Hasil pengujian pada sampel *Laughter* dengan skenario penyisipan sesudah *downmix* menunjukkan bahwa mekanisme 2 level hierarki secara konsisten lebih efisien dibandingkan mekanisme 1 level di hampir seluruh variasi *frame size*. Pada N=1024 (Gambar 4.10(a)), performa ketiga mekanisme tampak kompetitif, namun mekanisme 2 level mulai menunjukkan keunggulan dengan mencapai nilai $\log_{10}(\text{EP})$ sebesar -1.75 pada WE 6.4. Superioritas ini semakin jelas pada N=2048 (Gambar 4.10(b)), di mana mekanisme 2 level mencapai nilai $\log_{10}(\text{EP})$ sebesar -1.98 dengan WE 12.9, sedangkan mekanisme 1 level membutuhkan energi lebih besar (WE 16.0) untuk mencapai nilai $\log_{10}(\text{EP})$ sebesar -2.15. Pada ukuran *frame* menengah N=3072 dan N=4096, mekanisme 2 level mempertahankan dominasinya dengan kurva penurunan yang lebih landai. Pada N=4096 (Gambar 4.10(d)), mekanisme ini mencapai nilai $\log_{10}(\text{EP})$ sebesar -1.85 hanya dengan WE 12.4, lebih hemat dibanding mekanisme 1 level (WE 15.6). Tren ini dikukuhkan pada

$N=5120$ (Gambar 4.10(e)), di mana mekanisme 2 level mencapai nilai $\log_{10}(EP)$ sebesar -2.05 pada WE 15.6, sementara mekanisme 1 level memerlukan energi hingga 19.3. Secara keseluruhan, analisis pada sampel *Laughter* menyimpulkan bahwa struktur sinyal hasil *downmix* 2 level lebih kondusif bagi penyisipan data, terutama pada *frame* panjang, karena mampu mengonversi energi *watermark* menjadi penurunan tingkat kesalahan secara lebih efisien daripada mekanisme 1 level.

Berbeda dengan sampel audio sebelumnya (seperti *Drum* atau *Vivaldi*) di mana mekanisme 1 Level seringkali unggul pada *frame* besar, analisis pada *Laughter* menunjukkan sebaliknya. Mekanisme 3 Level Hierarki terbukti lebih efisien dan stabil, terutama pada ukuran *frame* menengah hingga besar (2048 - 5120). Pada konfigurasi ini, target *Error Probability* rendah (misalnya $\log_{10}(EP)$ sebesar -2 .) dapat dicapai dengan *Watermark Energy* yang lebih rendah dibandingkan mekanisme 1 Level. Hal ini mengindikasikan bahwa untuk sinyal dengan karakteristik ini, proses *downmix* bertingkat (2 level) justru menghasilkan sinyal mono yang lebih optimal sebagai wadah penyisipan data.

Secara keseluruhan, hasil pengujian pada kelima sampel audio menunjukkan bahwa mekanisme penyisipan pada sinyal *downmix*, baik yang melalui tiga level hierarki maupun yang disederhanakan menjadi dua dan satu level, mampu menghasilkan performa yang baik. Kelima audio memperlihatkan pola penurunan *Error Probability* yang konsisten ketika *Watermark Energy* ditingkatkan. Meskipun masing-masing sampel memiliki karakteristik sinyal yang berbeda, seluruh grafik menunjukkan bahwa sistem mampu menurunkan *Error Probability* hingga berada di bawah $\log_{10}(EP)$ sebesar -2 . Hal ini menandakan bahwa sinyal mono yang dihasilkan dari ketiga jenis mekanisme *downmix* tersebut sama-sama efektif berfungsi sebagai wadah penyisipan data, dengan tingkat keberhasilan ekstraksi yang tinggi.

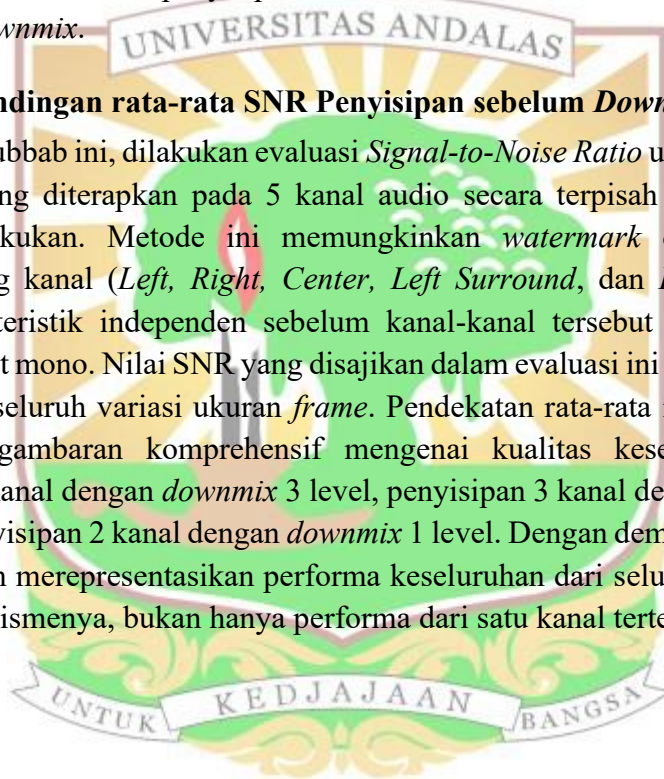
Mekanisme penyisipan pada audio *downmix* jauh lebih efisien secara energi dibandingkan sebelum *downmix* karena disisipkan langsung pada audio yang sudah melalui proses *downmixing*. Berbeda dengan mekanisme penyisipan sebelum *downmix*, energi *watermark* tersebar pada kanal yang disisipi pesan sehingga menghasilkan akumulasi energi yang lebih besar. Analisis terhadap variasi ukuran *frame* menyimpulkan bahwa mekanisme *downmix* 2 level cenderung memberikan stabilitas terbaik dan kebutuhan energi terendah pada ukuran *frame size* kecil (1024-2048). Namun, pada ukuran *frame* menengah hingga besar (3072-5120), mekanisme 1 level terbukti lebih baik dalam meminimalkan kebutuhan energi untuk mencapai *error probability* di bawah $\log_{10}(EP)$ sebesar -2 pada mayoritas sampel (*Drum*, *Vivaldi*, *Talking*, *Applause*). Dengan demikian, dapat disimpulkan bahwa penyederhanaan *downmix* menjadi 1 level sangat efektif untuk skenario berbasis *frame* besar guna penghematan energi, sedangkan mekanisme 3 dan 2 level lebih layak dipertahankan untuk menangani sinyal dengan penggunaan pada *frame* kecil.

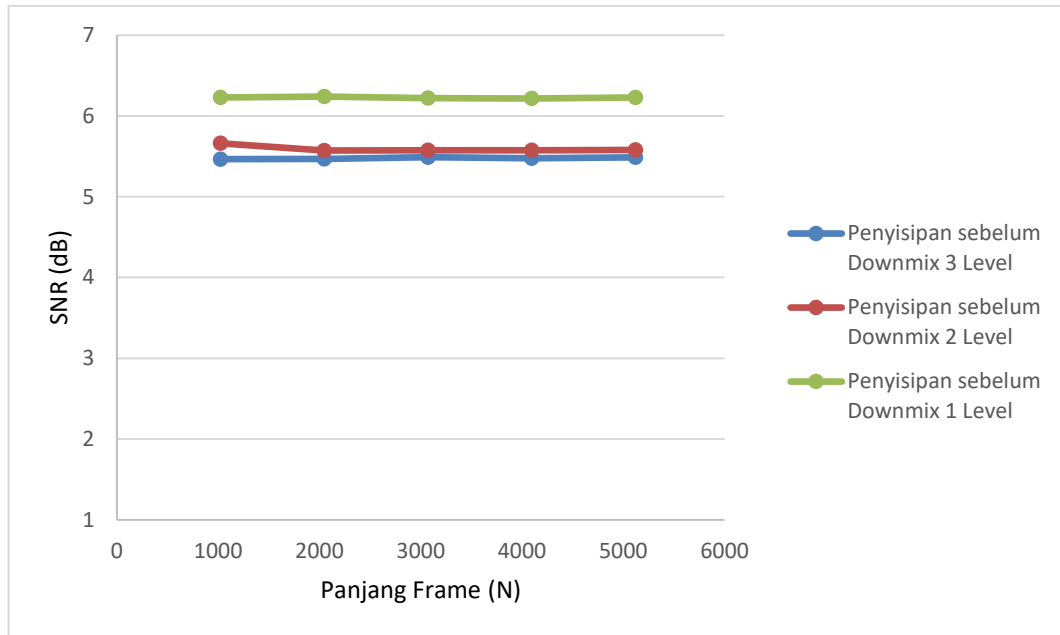
4.3. SNR (*Signal-to-Noise Ratio*)

Signal-to-Noise Ratio merupakan rasio perbandingan untuk mengukur sejauh mana kualitas sinyal hasil proses pengolahan masih mendekati sinyal aslinya. Nilai ini menggambarkan perbandingan antara energi sinyal asli terhadap energi *noise* atau gangguan yang muncul akibat proses modifikasi pada sinyal audio asli. Nilai SNR dinyatakan dalam satuan desibel (dB), dengan makna semakin besar nilai SNR, maka semakin kecil tingkat *noise* yang muncul, dan semakin baik pula kualitas sinyal hasil karena perbedaannya dengan sinyal asli semakin kecil. Sebaliknya, nilai SNR yang rendah menunjukkan bahwa sinyal hasil telah mengalami degradasi yang cukup besar akibat proses penyisipan atau kompresi. Pada hasil yang ditampilkan di bawah, nilai SNR diambil dari seluruh mekanisme penyisipan yaitu 3 variasi penyisipan sebelum *downmix* dan 3 variasi penyisipan pada audio *downmix*.

4.3.1. Perbandingan rata-rata SNR Penyisipan sebelum *Downmix*

Pada subbab ini, dilakukan evaluasi *Signal-to-Noise Ratio* untuk mekanisme penyisipan yang diterapkan pada 5 kanal audio secara terpisah sebelum proses *downmix* dilakukan. Metode ini memungkinkan *watermark* disisipkan pada masing-masing kanal (*Left*, *Right*, *Center*, *Left Surround*, dan *Right Surround*) dengan karakteristik independen sebelum kanal-kanal tersebut dikombinasikan menjadi format mono. Nilai SNR yang disajikan dalam evaluasi ini merupakan hasil rata-rata dari seluruh variasi ukuran *frame*. Pendekatan rata-rata ini dipilih untuk memberikan gambaran komprehensif mengenai kualitas keseluruhan sistem penyisipan 5 kanal dengan *downmix* 3 level, penyisipan 3 kanal dengan *downmix* 2 level, dan penyisipan 2 kanal dengan *downmix* 1 level. Dengan demikian, nilai SNR yang diperoleh merepresentasikan performa keseluruhan dari seluruh kanal sesuai dengan mekanismenya, bukan hanya performa dari satu kanal tertentu.



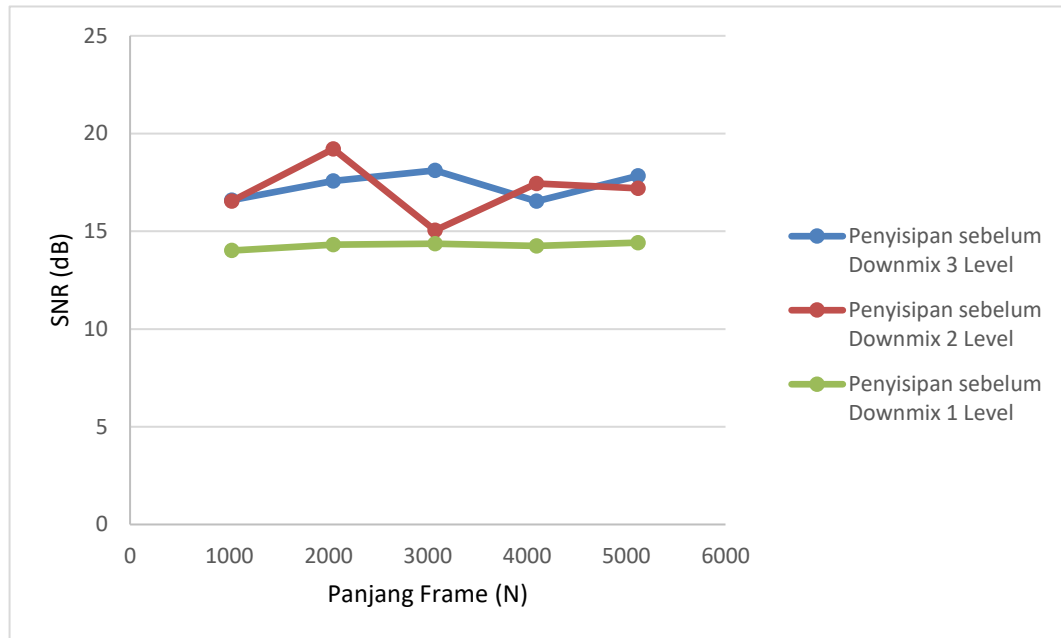


Gambar 4. 7 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan sebelum *Downmix* sampel audio *Applause*

Berdasarkan grafik pada Gambar 4.11, Visualisasi data perbandingan antara rata-rata *Signal-to-Noise Ratio* (SNR) di setiap ukuran *frame* sampel audio *Applause* cenderung bergerak landai dan berada di kisaran angka yang rendah untuk ketiga variasi mekanisme. Mekanisme penyisipan 2 kanal dengan *downmix* 1 level terlihat memiliki posisi grafik yang sedikit lebih tinggi dibandingkan mekanisme 3 kanal maupun 5 kanal. Seiring dengan kenaikan panjang *frame* dari 1024 hingga 5120, garis grafik perlahan-lahan merangkak naik, namun kenaikannya tidak terlalu signifikan. Pada panjang *frame* 1024, mekanisme 2 kanal mencatatkan nilai SNR di kisaran 6.1 dB, dan kurvanya bergerak naik secara perlahan namun stabil hingga mencapai sekitar 6.5 dB pada panjang *frame* 5120. Sebaliknya, mekanisme penyisipan 5 kanal (*downmix* 3 level) dan 3 kanal (*downmix* 2 level) terlihat mengalami stagnan, dengan kurva yang cenderung mendatar dan tertahan di nilai SNR yang lebih rendah, yaitu pada kisaran 5.4 dB hingga 5.6 dB sepanjang rentang *frame*.

Untuk menilai signifikansi fisik dari perbedaan yang terlihat kecil dalam satuan Desibel (dB) ini, dapat dilakukan konversi ke dalam bentuk skalar (linear) yang merepresentasikan perbandingan daya murni. Sebagai ilustrasi perhitungan pada ukuran *frame* 5120, mekanisme 2 kanal menghasilkan SNR sekitar 6.25 dB yang jika dikonversi ($10^{0.625}$) setara dengan nilai skalar 4.21, sementara mekanisme 5 kanal menghasilkan 5.50 dB ($10^{0.550}$) atau setara dengan nilai skalar 3.55. Nilai skalar ini menunjukkan berapa kali lipat kekuatan sinyal audio asli dibandingkan dengan noise *watermark*, yang artinya sinyal audio masing-masing adalah 4,21 kali dan 3,55 kali lebih kuat dari gangguan *noise*. Selisih skalar yang hanya sebesar 0,66 ini dalam konteks akustik dan pemrosesan sinyal digital

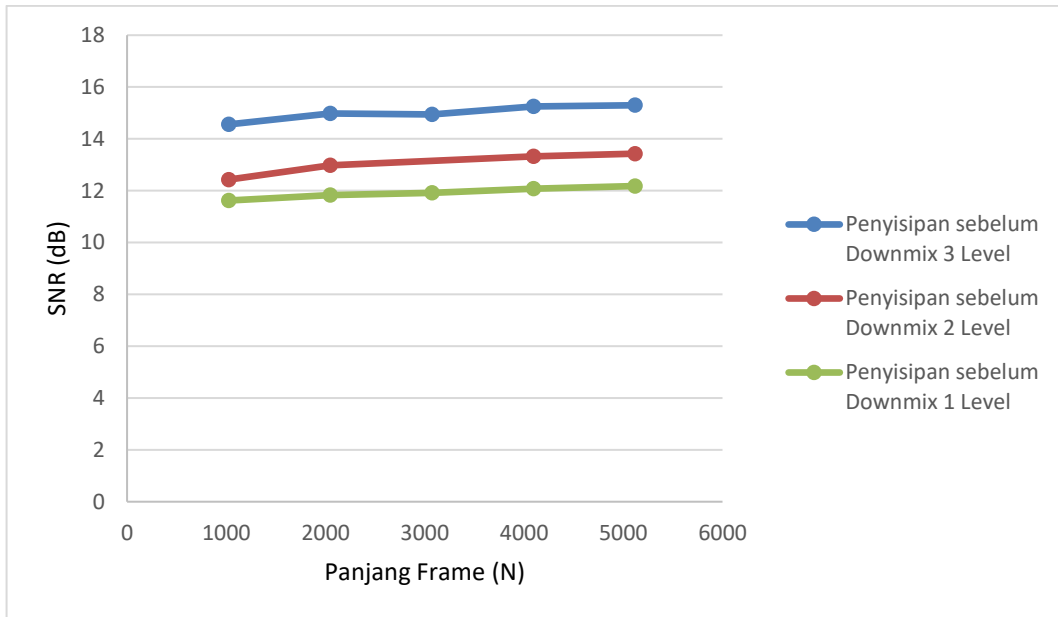
dianggap sangat kecil. Hal ini menegaskan bahwa secara fisik, tingkat dominasi sinyal audio terhadap *noise* yang dihasilkan oleh kedua mekanisme tersebut berada pada level yang nyaris identik, sehingga penambahan kompleksitas kanal pada tahap sebelum *downmix* tidak memberikan dampak perbedaan kualitas yang berarti secara statistik maupun persepsi.



Gambar 4. 8 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan sebelum *Downmix* sampel audio *Vivaldi*

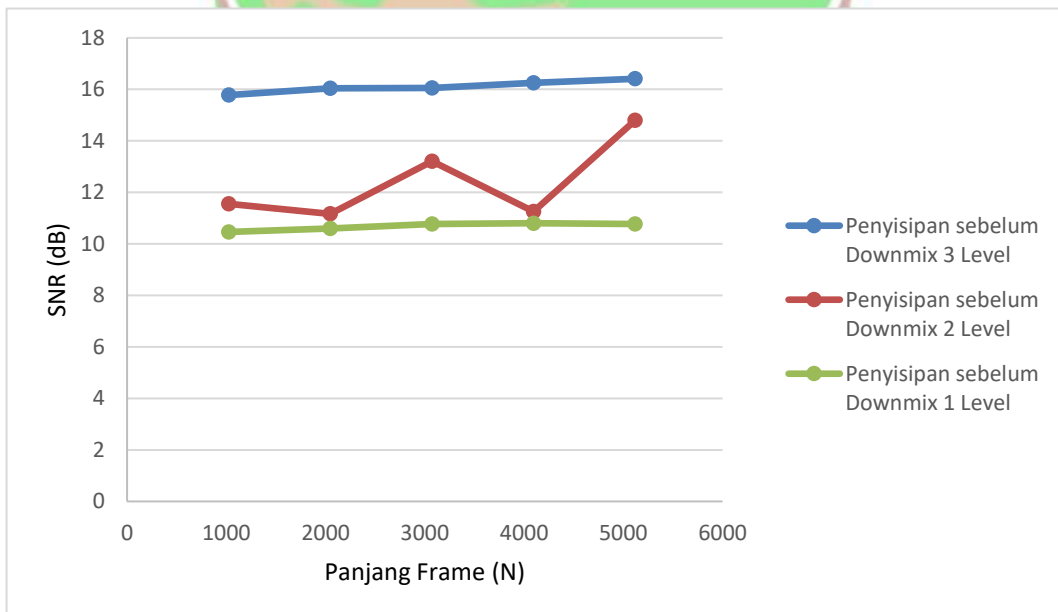
Pada sampel *Vivaldi*, Gambar 4.12 memperlihatkan tren mekanisme penyisipan 3 kanal dengan *downmix* 2 level muncul sebagai konfigurasi yang paling optimal. Grafik memperlihatkan kurva mekanisme 3 kanal yang lebih tinggi di atas yang lain, dimulai dari SNR sekitar 16.5 dB pada *frame* 1024 dan terus menanjak hingga berada di angka 17.5 dB pada *frame* 5120. Sementara itu, mekanisme 5 kanal dan 2 kanal berada di posisi yang lebih rendah dengan nilai SNR yang berhimpitan di kisaran 14 dB hingga 15 dB. Bentuk kurva pada sampel ini secara umum menunjukkan nilai yang tidak linear, ditandai dengan semakin besar ukuran *frame* yang digunakan, kemampuan sistem dalam menyembunyikan *watermark* di antara sampel audio *Vivaldi* tidak selalu membaik, terutama pada konfigurasi 3 kanal.

Pada sampel *Talking*, nilai SNR yang lebih tinggi beralih sepenuhnya ke mekanisme penyisipan 5 kanal dengan *downmix* 3 level. Grafik pada Gambar 4.13 memperlihatkan pola yang sangat rapi di mana ketiga garis bergerak naik secara linear dan hampir sejajar satu sama lain seiring penambahan panjang *frame*.



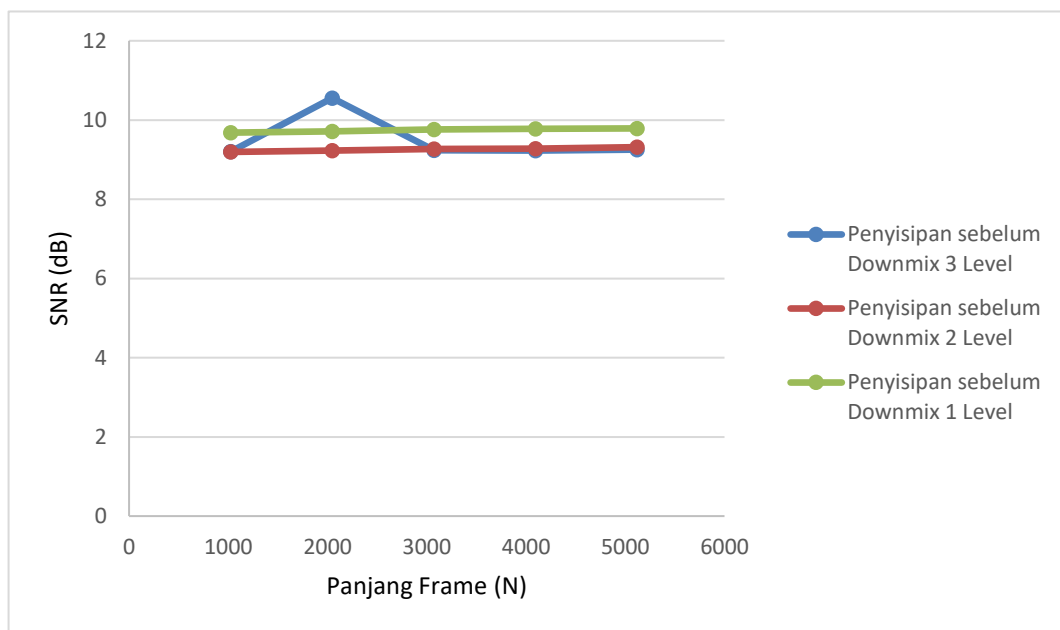
Gambar 4. 9 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan sebelum *Downmix* sampel audio *Talking*

Di bawahnya, mekanisme penyisipan 3 kanal mengikuti tren kenaikan yang sama, bergerak dari 12.4 dB menuju 13.4 dB. Mekanisme penyisipan 2 kanal berada di posisi terbawah, bergerak dari 11.6 dB naik secara stabil ke 12.2 dB. Tidak ada fluktuasi tajam pada sampel ini, seluruh mekanisme menunjukkan respons positif yang seragam terhadap kenaikan panjang *frame*, di mana semakin besar nilai *N*, semakin tinggi nilai SNR yang dihasilkan.



Gambar 4. 10 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan sebelum *Downmix* sampel audio *Drum*

Pada sampel *Drum*, Gambar 4.14 memperlihatkan pola yang mirip dengan SNR audio *Talking*, di mana mekanisme penyisipan 5 kanal kembali menunjukkan nilai SNR tertinggi. Mekanisme penyisipan 5 kanal dengan *downmix* 3 level memimpin jauh di atas dengan kestabilan tinggi, bergerak lurus dari 15.8 dB naik sedikit demi sedikit hingga 16.4 dB. Sebaliknya, mekanisme penyisipan 2 kanal berada di posisi paling bawah dengan garis datar di kisaran 10.4 dB hingga 10.8 dB. Namun, mekanisme penyisipan 3 kanal (garis merah) menunjukkan perilaku yang kurang stabil, dimulai dari 11.5 dB, turun sedikit, lalu melonjak naik ke 13.2 dB pada N=3072, jatuh kembali ke 11.2 dB pada N=4096, dan akhirnya melonjak drastis secara signifikan hingga hampir menyentuh 14.8 dB pada panjang *frame* 5120.



Gambar 4. 11 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan sebelum *Downmix* sampel audio *Laughter*

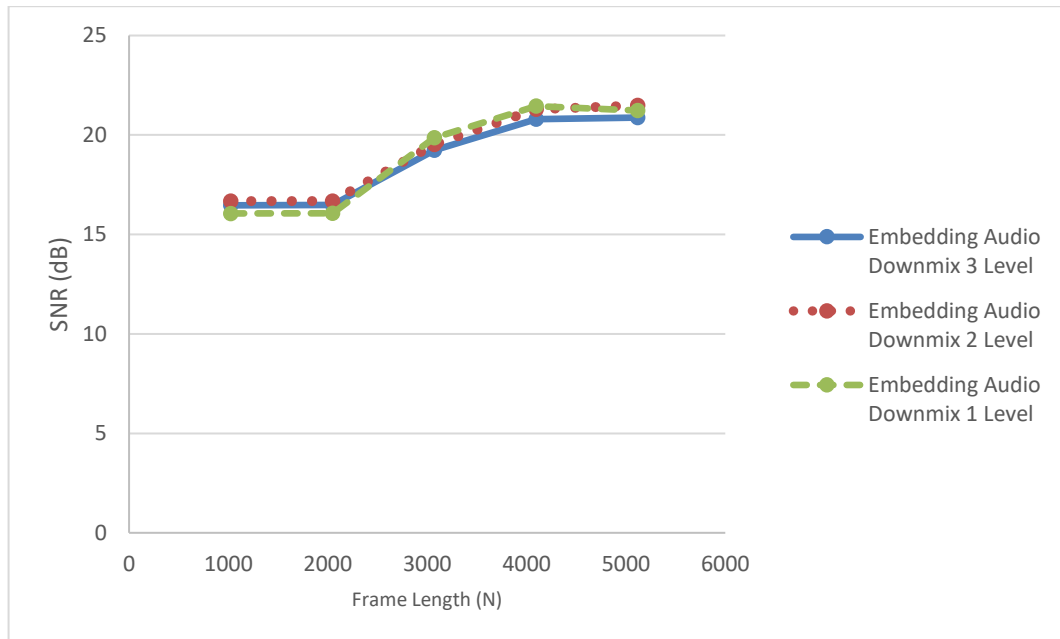
Pada audio *Laughter*, Gambar 4.15 memperlihatkan adanya fenomena unik pada mekanisme penyisipan 5 kanal (garis biru). Meskipun pada titik awal (N=1024) dan titik-titik akhir (N=3072 ke atas) nilainya sangat rendah di kisaran 9.2 dB, mekanisme ini mengalami lonjakan pada N=2048 yang mencapai angka 10.55 dB, sebelum akhirnya turun kembali. Sebaliknya, mekanisme penyisipan 2 kanal (garis hijau) justru menjadi yang paling unggul secara rata-rata, bergerak stabil dari 9.68 dB dan naik perlahan hingga 9.79 dB tanpa gangguan fluktuasi. Sementara itu, mekanisme penyisipan 3 kanal (garis merah) konsisten berada di posisi terendah dengan pergerakan yang sangat datar dan kenaikan minimal dari 9.19 dB ke 9.31 dB sepanjang sumbu x. Secara keseluruhan untuk *Laughter*, hanya garis biru yang menunjukkan ketidakstabilan pada panjang *frame* tertentu, sementara dua mekanisme lainnya cenderung linear.

Secara keseluruhan, evaluasi terhadap kelima sampel audio menunjukkan bahwa respon nilai SNR terhadap hierarki mekanisme penyisipan, mulai dari 5 kanal (*downmix* 3 level), 3 kanal (*downmix* 2 level), hingga 2 kanal (*downmix* 1 level), tidak memiliki pola linear yang seragam untuk semua jenis audio. Mekanisme dengan kompleksitas tertinggi, yakni penyisipan 5 kanal, terbukti unggul dan menghasilkan SNR tertinggi pada sampel audio *Talking* dan *Drum*, serta bersaing ketat pada audio *Vivaldi*. Sebaliknya, pada sampel audio *Applause* dan *Laughter*, mekanisme yang paling sederhana yakni penyisipan 2 kanal justru menunjukkan performa yang lebih stabil dan unggul dibandingkan mekanisme dengan level *downmix* yang lebih tinggi.

Dalam hal ini, sinyal audio dari kelima kanal ketika dilakukan proses *downmix* atau penjumlahan sinyal, menghasilkan penguatan energi yang bersifat konstruktif secara kuadrat. Artinya, ketika lima kanal audio digabungkan pada mekanisme *downmix* 3 level, lonjakan amplitudo gabungan setiap kanal saling mendukung. Sebaliknya, sinyal *watermark* dibentuk dari bit pesan acak (+1 dan -1) yang bersifat inkoheren atau tidak memiliki korelasi antar-kanal. Akibatnya, saat *downmix* terjadi, energi *watermark* tidak bertambah secepat energi audio karena adanya fenomena interferensi destruktif, di mana bit positif dari satu kanal seringkali bertemu dengan bit negatif dari kanal lain dan saling meniadakan (*cancellation*). Kondisi inilah yang menyebabkan nilai SNR pada mekanisme 5 kanal terlihat lebih besar dibandingkan mekanisme 2 kanal.

Meskipun terdapat variasi performa antar mekanisme, secara umum kualitas sinyal yang dihasilkan dapat dikatakan belum terlalu baik atau kurang optimal. Indikator utamanya adalah masih banyaknya ditemukan nilai SNR yang berada di bawah angka 10 dB, sebuah ambang batas yang biasanya menandakan adanya gangguan *noise* yang cukup dominan dibandingkan sinyal aslinya. Kondisi nilai rendah ini terlihat sangat konsisten pada seluruh percobaan sampel audio *Applause* dan mayoritas pada sampel audio *Laughter* di berbagai mekanisme. Rendahnya capaian SNR ini menunjukkan bahwa proses penyisipan sebelum *downmix* MPEG *Surround* yang diterapkan masih menyisakan distorsi yang signifikan, sehingga integritas audio pasca-proses masih memerlukan perbaikan lebih lanjut untuk mencapai standar kualitas yang diharapkan.

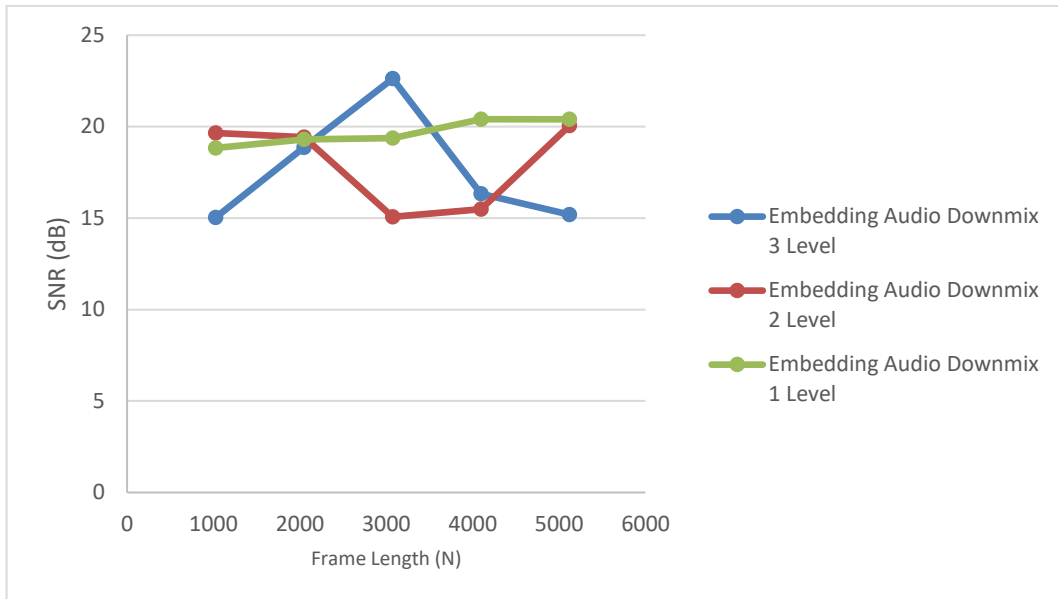
4.3.2. Perbandingan rata-rata SNR Penyisipan Audio *Downmix*



Gambar 4.12 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan sesudah *Downmix* sampel audio *Applause*

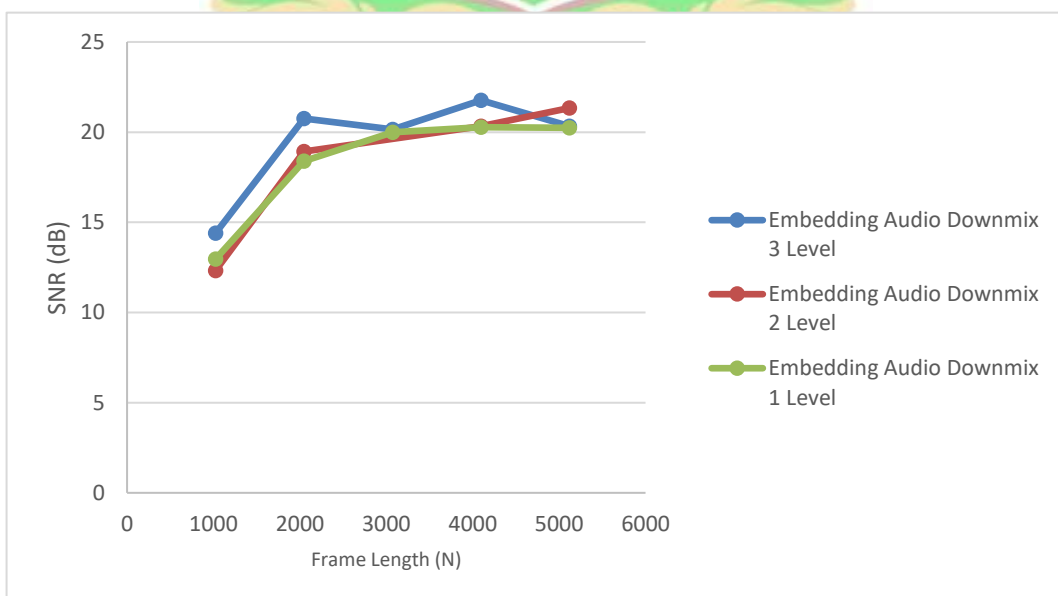
Perbandingan nilai rata-rata SNR pada mekanisme penyisipan audio *downmix* sampel *Applause* ditampilkan pada Gambar 4.16 diatas. Pada grafik ini, terlihat pola pergerakan yang sangat harmonis di mana ketiga garis mekanisme, baik *downmix* 3 level, 2 level, maupun 1 level, bergerak beriringan dengan sangat rapat, bahkan hampir saling menumpuk. Ketiga mekanisme memulai nilai SNR di kisaran 16 dB hingga 16.5 dB pada panjang *frame* (N) 1024. Grafik kemudian menunjukkan tren kenaikan yang signifikan mulai dari N=2048 hingga N=3072, melonjak dari kisaran 16 dB menuju angka 20 dB. Setelah lonjakan tersebut, kurva melandai dan cenderung stabil di angka yang cukup tinggi, yaitu sekitar 21 dB hingga 21.5 dB pada N=5120. Secara visual, nilai SNR meningkat konsisten dan mencapai titik jenuh yang tinggi seiring bertambahnya panjang *frame*. Pada mekanisme sebelumnya, audio *Applause* hanya mampu menghasilkan SNR di kisaran 5 dB hingga 6 dB dengan garis yang datar. Sebaliknya, mekanisme penyisipan setelah *downmix* ini mampu mendongkrak nilai SNR hingga menembus angka 20 dB.

Beralih ke audio *Vivaldi*, grafik Gambar 4.17 menunjukkan dinamika yang sangat kontras antar mekanisme. Mekanisme penyisipan pada *downmix* 1 level (garis hijau) terlihat stabil, bergerak lurus dengan sedikit kenaikan dari 19 dB menuju 20.5 dB. Sebaliknya, dua mekanisme lainnya sangat fluktuatif dan saling berlawanan arah. Mekanisme *downmix* 3 level (garis biru) memulai dari 15 dB, melonjak hingga mencapai puncak tertinggi di 22.5 dB pada N=3072, lalu turun kembali secara drastis ke angka 15 dB di akhir grafik.



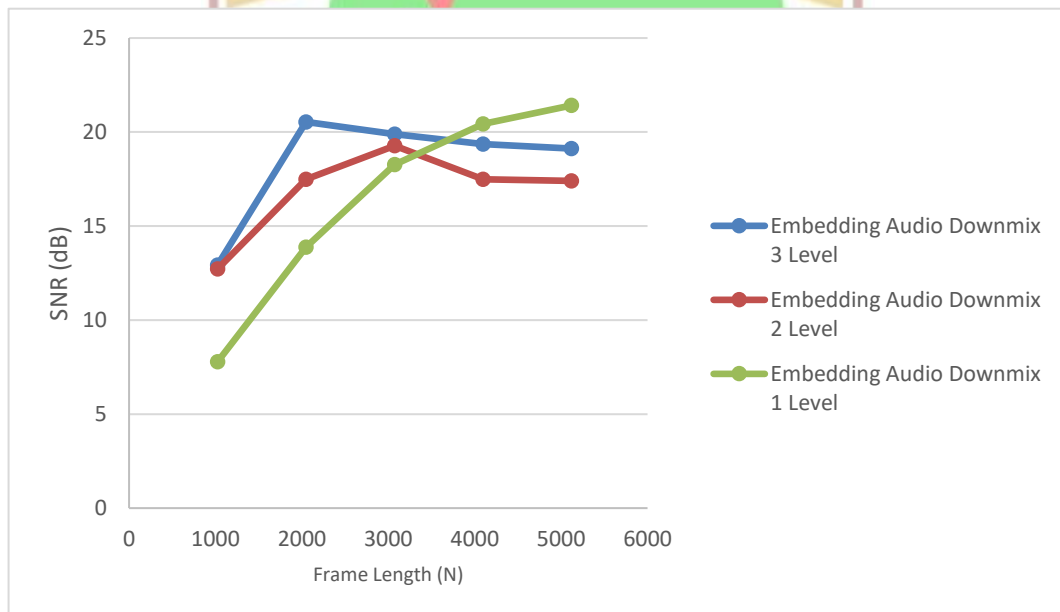
Gambar 4.13 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan Audio *Downmix* sampel audio *Vivaldi*

Sementara itu, mekanisme *downmix* 2 level (garis merah) justru mengalami penurunan pada titik tengah, dimulai dari 19.5 dB, turun ke 15 dB pada $N=3072$, sebelum akhirnya naik kembali menyentuh angka 20 dB pada panjang *frame* maksimum. Dibandingkan dengan metode sebelumnya, mekanisme penyisipan setelah *downmix* ini mencapai nilai puncak yang lebih tinggi. Sebelumnya, nilai tertinggi *Vivaldi* hanya menyentuh 19.2 dB dengan banyak fluktuasi di angka 14-16 dB. Pada metode ini, stabilitas pada *downmix* 1 level jauh lebih baik dengan nilai yang konsisten berada di atas 18 dB, serta nilai puncak pada *downmix* 3 level mampu menembus angka 22 dB.



Gambar 4. 14 Nilai Perbandingan rata-rata SNR dengan Panjang *Frame* Mekanisme Penyisipan Audio *Downmix* sampel audio *Talking*

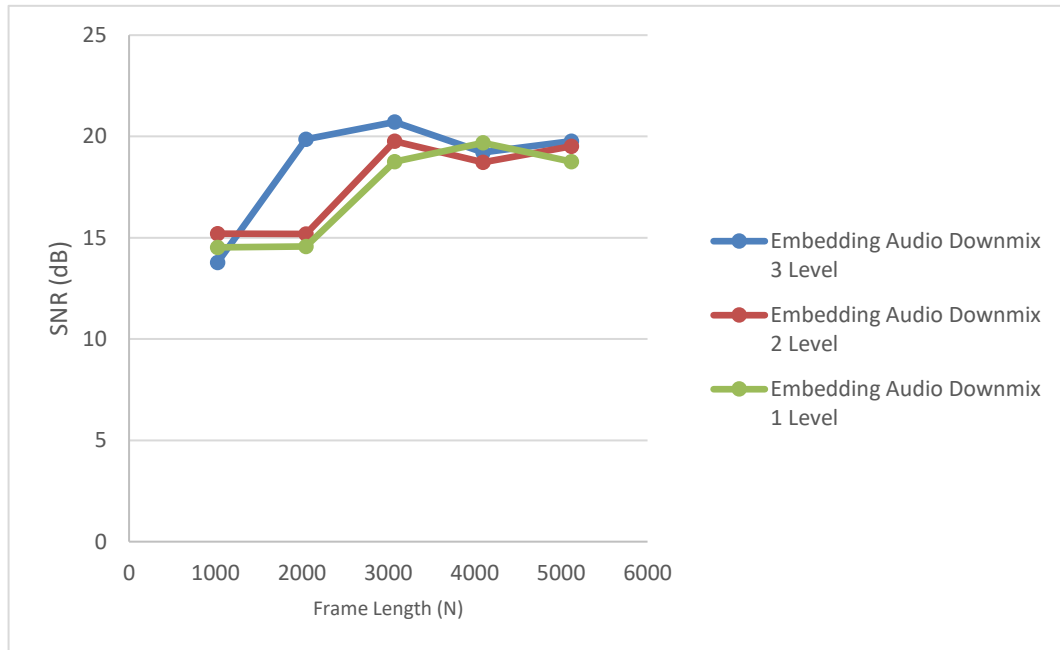
Pada grafik audio *Talking*, ketiga garis mekanisme menunjukkan tren positif dengan menaik dari kiri bawah ke kanan atas. Gambar 4.58 menunjukkan mekanisme *downmix* 3 level memimpin di awal dengan nilai 14.5 dB dan dengan cepat naik ke level 20.5 dB pada $N=2048$, lalu bertahan stabil di kisaran tersebut hingga akhir. Mekanisme *downmix* 2 level dan 1 level memulai dari titik yang lebih rendah, yaitu sekitar 12 dB hingga 13 dB. Keduanya kemudian merangkak naik secara bertahap mengejar ketinggalan, di mana pada panjang *frame* 5120, mekanisme *downmix* 2 level (garis merah) justru berhasil menyalip dan mencatatkan nilai tertinggi di angka 21.5 dB, sedikit di atas mekanisme lainnya yang berkumpul di angka 20.5 dB. Kurva secara keseluruhan menggambarkan konvergensi, di mana ketiga garis yang terpisah di awal akhirnya bertemu di rentang nilai yang sama di akhir. Pada penyisipan sebelum *downmix*, audio *Talking* hanya bergerak naik perlahan dari 11 dB dan mencapai nilai tertinggi di angka 15.3 dB. Dengan mengubah urutan menjadi penyisipan setelah *downmix*, nilai SNR berhasil ditingkatkan secara signifikan hingga mencapai rata-rata di atas 20 dB pada panjang *frame* yang besar.



Gambar 4. 15 Nilai Perbandingan rata-rata SNR dengan Panjang *Frame* Mekanisme Penyisipan Audio *Downmix* sampel audio *Drum*

Pada sampel audio *Drum*, Gambar 4.19 memperlihatkan fenomena di mana mekanisme *downmix* 1 level (garis hijau) memulai dari nilai terendah di 7.8 dB, namun menaik drastis secara linear hingga menyalip dua mekanisme lainnya dan berakhir di posisi tertinggi yaitu 21.3 dB. Sementara itu, mekanisme *downmix* 3 level dan 2 level bergerak lebih stabil di kisaran tinggi, mulai dari 12-13 dB dan

bertahan di area 17 dB hingga 19 dB. Jika dikomparasi dengan metode sebelumnya, peningkatan kualitas juga terlihat jelas. Pada mekanisme penyisipan sebelum *downmix*, nilai maksimum yang bisa dicapai audio *Drum* hanya sekitar 16.4 dB, sedangkan pada mekanisme ini, nilai SNR mampu mencapai angka 21.3 dB, terutama pada mekanisme *downmix* 1 level yang menunjukkan respons sangat baik terhadap penambahan panjang *frame*.



Gambar 4. 16 Nilai Perbandingan rata-rata SNR dengan Panjang Frame Mekanisme Penyisipan Audio *Downmix* sampel audio *Laughter*

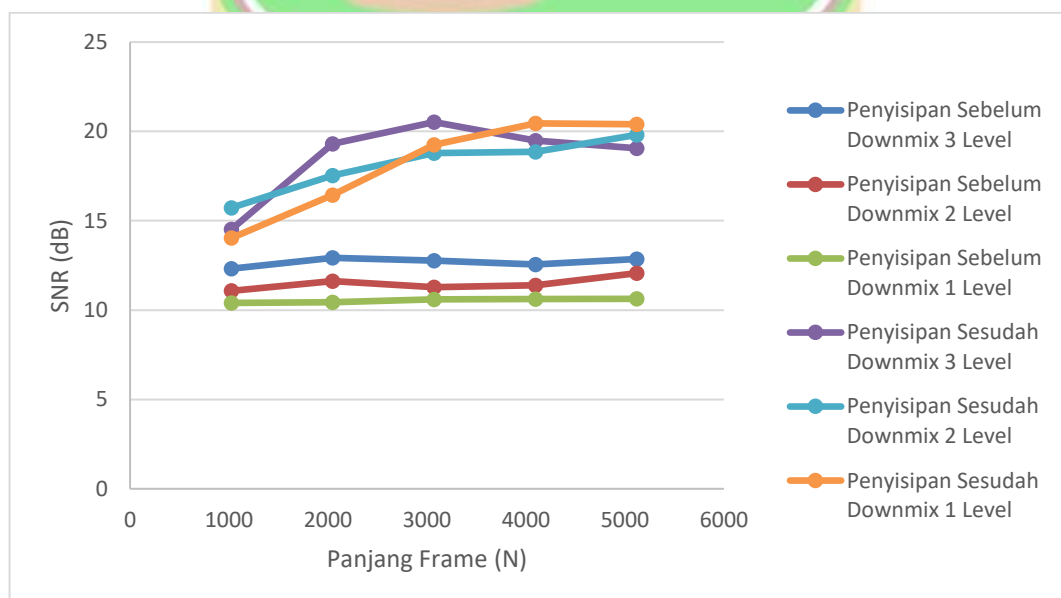
Pada sampel audio terakhir yakni *Laughter*, grafik Gambar 4.20 menunjukkan kurva yang melengkung naik dan kemudian stabil. Ketiga mekanisme memulai di titik yang rapat sekitar 13.8 dB hingga 15 dB. Terjadi kenaikan signifikan pada $N=2048$ dan $N=3072$ di mana nilai SNR mencapai puncaknya di kisaran 19 dB hingga 20.7 dB, sebelum sedikit menurun dan stabil di angka 18.8 dB hingga 19.7 dB. Ini merupakan perbaikan yang sangat masif dibandingkan mekanisme sebelumnya. Pada penyisipan sebelum *downmix*, audio *Laughter* memiliki nilai SNR yang sangat buruk, mayoritas di bawah 10 dB. Dengan metode penyisipan setelah *downmix* ini, nilai SNR melonjak dua kali lipat lebih, menjadikannya jauh lebih layak dan berkualitas dengan nilai rata-rata yang konsisten mendekati 20 dB.

Secara keseluruhan, hasil pengujian pada mekanisme penyisipan yang dilakukan sesudah *downmix*, baik hierarki 3 level, 2 level, maupun 1 level, menunjukkan peningkatan performa yang sangat positif dan konsisten di seluruh sampel audio. Mayoritas grafik menampilkan tren kenaikan nilai SNR yang stabil seiring dengan bertambahnya panjang *frame* (N), di mana nilai rata-rata yang dihasilkan mampu bertahan di angka yang tinggi, sering kali menembus kisaran 20

dB hingga 22 dB pada titik optimalnya. Stabilitas ini menandakan bahwa mekanisme penyisipan setelah proses *downmix* jauh lebih bersahabat terhadap kualitas sinyal audio, meminimalkan gangguan *noise*, dan menghasilkan *output* yang lebih jernih dibandingkan percobaan sebelumnya.

Perbandingan antara mekanisme ini dengan mekanisme sebelumnya (penyisipan sebelum *downmix*) memperlihatkan perbedaan kualitas yang sedikit mencolok. Sampel audio *Applause* dan *Laughter*, yang pada mekanisme sebelumnya terpuruk dengan nilai SNR sangat rendah di bawah 10 dB, kini mengalami lonjakan kualitas drastis hingga mencapai kisaran 19 dB sampai 21 dB pada mekanisme penyisipan sesudah *downmix*. Jika pada metode sebelumnya banyak ditemukan garis grafik yang datar di level rendah, pada metode ini justru menunjukkan respons dinamis yang terus membaik, membuktikan bahwa integritas sinyal jauh lebih terjaga ketika proses *downmix* diselesaikan terlebih dahulu. Sebagai kesimpulan akhir, dapat dipastikan bahwa urutan proses dalam sistem ini memegang peranan kunci terhadap nilai SNR akhir. Strategi melakukan penyisipan pada audio yang sudah di-*downmix* terbukti jauh lebih baik dibandingkan melakukan penyisipan sebelum *downmix*. Rendahnya nilai SNR pada mekanisme penyisipan sebelum *downmix* mengindikasikan bahwa proses *downmix* MPEG Surround bersifat destruktif terhadap data yang telah disisipkan sebelumnya, sehingga daya *noise* atau distorsi menjadi cukup dominan dan mengganggu daya sinyal asli, yang berimplikasi pada penurunan kualitas perseptual (*imperceptibility*) audio secara signifikan.

4.3.3. Perbandingan rata-rata SNR seluruh Mekanisme Penyisipan dan Level Hierarki Sistem *Downmix* MPEG Surround



Gambar 4. 17 Grafik Rata-rata SNR seluruh Mekanisme Penyisipan dan Level Hierarki Sistem *Downmix* MPEG Surround

Grafik pada Gambar 4.21 diatas memperlihatkan rangkuman nilai SNR rata-rata dari seluruh sampel audio, di mana terlihat perbedaan yang sangat mencolok antara kelompok mekanisme penyisipan sebelum *downmix* (garis-garis bawah) dengan mekanisme penyisipan pada audio *downmix* (garis-garis atas). Pada kelompok pertama yaitu mekanisme penyisipan sebelum *downmix*, ketiga variasi garis menunjukkan tren yang sangat datar dan bernilai rendah. Garis biru tua (Penyisipan 5 Kanal, *Downmix* 3 Level) bergerak statis di kisaran 12.3 dB hingga 12.9 dB. Di bawahnya, garis merah (Penyisipan 3 Kanal, *Downmix* 2 Level) berada di kisaran 11 dB hingga 12 dB, sementara garis hijau (Penyisipan 2 Kanal, *Downmix* 1 Level) menjadi yang terendah dengan nilai konstan yang nyaris tidak bergerak di angka 10.5 dB sepanjang sumbu x. Hal ini menunjukkan bahwa pada metode ini, perubahan panjang *frame* (*frame length*) tidak memberikan dampak signifikan terhadap perbandingan antara sinyal asli dengan *noise*.

Sebaliknya, kelompok mekanisme penyisipan sesudah *downmix* menunjukkan dinamika yang jauh lebih positif dengan tren kenaikan yang signifikan seiring bertambahnya panjang *frame*. Garis ungu (Penyisipan Audio *Downmix* 3 Level) memulai dari 14.5 dB pada N=1024 dan mengalami peningkatan hingga mencapai puncaknya di 20.4 dB pada N=3072. Pola serupa diikuti oleh garis biru muda (Penyisipan sesudah *Downmix* 2 Level) yang bergerak dari 15.2 dB naik menuju 19.8 dB. Garis oranye (Penyisipan sesudah *Downmix* 1 Level) menunjukkan peningkatan paling stabil, dimulai dari 14 dB dan terus naik secara linear hingga menyentuh angka tertinggi di 20.3 dB pada akhir grafik. Ketiga garis ini membuktikan bahwa metode penyisipan setelah *downmix* sangat responsif terhadap penambahan ukuran *frame* untuk meningkatkan rasio perbandingan sinyal asli dengan *noise*.

Sebagai kesimpulan perbandingan, terdapat kesenjangan nilai SNR yang cukup jauh antara kedua pendekatan tersebut. Mekanisme penyisipan sebelum *downmix* memiliki rata-rata nilai SNR yang tertahan di bawah 13 dB, yang berarti bahwa kualitas sinyal kurang baik akibat penyisipan pesan rahasia yang dilakukan pada lebih dari 1 kanal serta proses *downmix* menghasilkan distorsi yang besar. Sinyal *watermark* yang disisipkan pada masing-masing kanal audio asli melewati serangkaian proses *downmix*, baik 3 tingkat, 2 tingkat, ataupun hanya 1 tingkat, sehingga sinyal *watermark* menjadi rusak atau “kotor” karena terjadi penggabungan dari 2 kanal menjadi 1 kanal dan hal tersebut terjadi berulang. Sementara itu, mekanisme penyisipan sesudah *downmix* mampu menghasilkan rata-rata SNR yang jauh lebih tinggi, dengan selisih keunggulan mencapai 7 dB hingga 8 dB pada *frame length* yang lebih besar (N=3072 ke atas), menjadikan metode ini jauh lebih efektif dalam mempertahankan daya sinyal audio terhadap sinyal *noise*. Mekanisme ini menyisipkan *watermark* langsung pada hasil akhir (*monomix*), sehingga *watermark energy* yang digunakan untuk mencapai target deteksi yang maksimal membutuhkan lebih sedikit sehingga sinyal audio asli jauh lebih dominan dibanding *noise* nya dibanding mekanisme penyisipan sebelum *downmix*.

Lebih lanjut, analisis terhadap rentang rata-rata nilai SNR yang diperoleh dari seluruh mekanisme pengujian, yakni berkisar antara 11 dB hingga 20 dB, menjadi dasar pertimbangan teknis ditiadakannya pengukuran kualitas menggunakan parameter *Objective Difference Grade* (ODG). Secara fundamental, algoritma pengukuran ODG (PEAQ) didesain untuk mendeteksi degradasi audio yang sangat halus pada sinyal dengan fidelitas tinggi, di mana distorsi sulit dibedakan oleh telinga manusia. Namun, rentang SNR di bawah 20 dB mengindikasikan bahwa sinyal audio stego telah mengalami distorsi yang sangat parah dengan level *noise* yang dominan, sehingga secara perseptual perbedaan antara audio asli dan stego sudah sangat nyata (*audible*) dan jauh melampaui ambang batas imperseptibilitas yang menjadi domain kerja algoritma ODG. Oleh karena itu, penggunaan ODG pada kondisi sinyal yang terdistorsi berat ini menjadi tidak valid, karena hasil yang diperoleh tidak akan lagi merepresentasikan gradasi kualitas persepsi manusia secara akurat.

