

PERBANDINGAN STEGANALISIS SINYAL WICARA BERFORMAT .WAV ANTARA METODE ANALISIS *CEPSTRAL* DAN *MEL-FREQUENCY CEPSTRAL COEFFICIENT* (MFCC)

Kevin Putra Dirgantoro¹, Bambang Hidayat², Nur Andini³

^{1, 2, 3}Fakultas Teknik Elektro, Universitas Telkom

¹kevputrad@telkomuniversity.ac.id,²bhidayat@telkomuniversity.ac.id,

³nurandini@telkomuniversity.ac.id

Abstrak

Teknik menyembunyikan pesan rahasia ke dalam suatu data tertentu atau yang biasa dikenal dengan steganografi mengalami perkembangan yang sangat pesat. Namun, ternyata metode penyembunyian pesan ini juga menimbulkan masalah, di antaranya pihak-pihak yang tidak bertanggung jawab menggunakan teknik tersebut untuk kegiatan kriminalitas. Oleh karena itu, diperlukan teknik untuk mendeteksi pesan tersembunyi di dalam suatu data. Teknik tersebut dikenal dengan istilah steganalisis. Pada penelitian ini, dilakukan analisis terhadap berkas sinyal wicara yang berformat .wav, dengan menggunakan dua metode, yaitu analisis *cepstral* dan *Mel-Frequency Cepstral Coefficient* (MFCC). Perbandingan dari kedua metode ini dilakukan untuk mengetahui metode mana yang lebih baik untuk mendeteksi data memiliki pesan rahasia atau tidak. Nilai akurasi yang didapat dengan menggunakan 45 data latih dan uji untuk metode analisis *cepstral* yaitu sebesar 51,11%, sedangkan untuk MFCC sebesar 77,78%. Nilai akurasi tersebut didapat dari ciri statistik yang terdiri dari nilai *kurtosis*, *skewness*, dan *standard deviation* yang dihasilkan dari kedua metode, dengan menggunakan metode klasifikasi *Support Vector Machine* (SVM).

Kata Kunci: analisis *cepstral*, *Mel-Frequency Cepstral Coefficient* (MFCC), steganalisis WAV

Abstract

A technique to hide secret message into a specific data or known as steganography, having a rapid progress. Unfortunately, this phenomena also triggers new problem: many of irresponsible parties use this method for criminal activity. Therefore, a technique for detecting a hidden message in a specific data is necessary. This technique is known as steganalysis. In this research, an analysis of speech signal file with .wav format was performed using two different methods: analysis *cepstral* and *Mel-Frequency Cepstral Coefficient* (MFCC). The methods were compared in order to know which method is better for hidden message detection. The accuracy value for *cepstral* analysis that using 45 train and test file is 51,11%, meanwhile for MFCC is 77,78%. The accuracy values were obtained from statistical features, that consist of *kurtosis*, *skewness*, and *standard deviation* from both methods, using *Support Vector Machine* (SVM) for its classification.

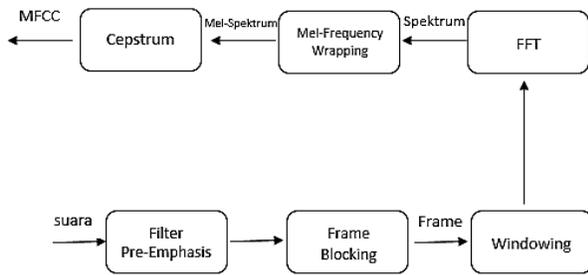
Key Words: *cepstral* analysis, MFCC, statistics, steganalysis, SVM, .WAV

1. Pendahuluan

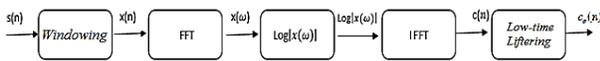
Perkembangan teknologi telekomunikasi semakin maju seiring dengan berjalannya waktu. Karena adanya telekomunikasi, hubungan komunikasi masih dapat berlangsung walaupun terpisahkan oleh jarak. Dalam berkomunikasi jarak jauh, terkadang ada pesan terkirim yang dapat diakses publik, namun ada juga yang bersifat rahasia demi menjaga privasi. Pengiriman pesan secara rahasia ini telah menjadi suatu kesenian yang memang tidak semua orang dapat melakukannya, yang disebut dengan Steganografi. Pesan rahasia yang dikirim melalui

cara steganografi dapat bermacam-macam jenisnya, misalkan, diubah menjadi bahasa yang sulit dimengerti, menggunakan mesin (salah satunya mesin enigma) atau dengan menyisipkan suatu pesan rahasia pada suatu media, baik gambar, audio bahkan video. [1]

Steganografi sangat dibutuhkan untuk mengirimkan pesan yang sangat rahasia, misalnya antar kepala negara yang bertukar informasi rahasia dengan media gambar. Namun, bagaimana jika steganografi ini dilakukan oleh orang yang tidak bertanggung jawab? Pengembangan teknik untuk memecahkan atau membongkar informasi yang telah disisipkan pada pesan rahasia menggunakan



Gambar 1. Blok Diagram MFCC



Gambar 2. Blok Diagram Analisis Cepstral

teknik steganografi juga diperlukan, yang disebut dengan steganalisis.[2]

Pada penelitian ini, analisis perbandingan akurasi steganalisis menggunakan dua metode dikaji. Media yang digunakan adalah sinyal wicara (*speech*), karena sinyal ini merupakan media yang sering disisipkan oleh pesan rahasia. Sedangkan format berkas *speech* yang digunakan adalah *.wav*, karena sering dipakai kebanyakan orang, memiliki struktur yang sederhana dan kualitas suaranya yang baik [3]. Metode yang dibandingkan yaitu metode analisis *Cepstral* dan *Mel-Frequency Cepstral Coefficient* (MFCC), dengan pertimbangan keduanya memiliki mekanisme kerja yang relatif sama, sehingga patut diteliti lebih jauh metode mana yang lebih efisien dan memiliki kualitas yang lebih baik. Nilai akurasi yang didapat merupakan nilai *root mean square* (rms) dari setiap *frame* sinyal wicara yang menghasilkan distribusi *error* yang dapat dicirikan dengan kurtosis, *skewness*, dan *standard deviation*.

Sinisuka dkk, telah dilakukan steganalisis menggunakan metode statistik *Mel-Frequency Cepstral Coefficient* (MFCC) yang menggunakan klasifikasi *Support Vector Machine* (SVM)[4]. Hasil akurasi yang didapatkan berdasarkan jumlah sisipan, yaitu sebesar 43,75% untuk sisipan yang sedikit, sedangkan untuk sisipan yang banyak nilai akurasinya mencapai 50%.

Gambar 1 menjelaskan bagaimana cara kerja metode MFCC untuk mendapatkan ekstraksi ciri MFCC. Hal pertama yang dilakukan adalah melakukan *pre-emphasis* menggunakan orde pertama *Finite Impuls Response* (FIR) dan *high pass filter* (HPF), yang didasari oleh hubungan antara masukan dan keluaran dalam domain waktu [5]. Kemudian dilakukan *frame blocking* dengan menggunakan sistem *overlap*, yang bertujuan untuk meminimalisir hilangnya ekstraksi ciri [6]. Setelah sinyal dibuat dalam bentuk *frame*, setiap *frame* dikalikan dengan *window hanning*, untuk menghindari terjadinya *discontinuities* [5]. Kemudian dilakukan *Fast Fourier*

Transform (FFT), untuk mengubah sinyal dari domain waktu ke domain frekuensi.

Pada proses *Mel-Frequency wrapping* ini, *filterbank* yang digunakan terdiri dari sepasang *Band Pass Filter* (BPF), yang kondisi *bandwidth* dan *spacing*-nya mendekati *critical band*-nya. Selain itu juga, bagian tengah dari *frequency filter* tersebut digunakan untuk meng-cover frekuensi-frekuensi penting untuk persepsi sinyal wicara [5]. Pada tahap terakhir, hasil mel spektrum dikonversikan kembali ke dalam domain waktu menggunakan invers FFT dengan fungsi *Discrete Cosine Transform* (DCT). Hasil akhir dari perhitungan tersebut dinamakan *Mel-Frequency Cepstral Coefficient* (MFCC), yang merupakan representasi dari spektrum sinyal wicara [7].

Telah dilakukan juga penelitian yang menggunakan metode analisis *Cepstral* oleh Rao dkk, namun penelitian yang dilakukan berfokus pada pendekatan peningkatan kata bicara dua tahap berbasis data yang pada *stage* kedua sisa dari *noise* di-suppress dan didapatkan peningkatan kualitas suara 20%, pada kondisi *noise* tertentu [8].

Gambar 2 merupakan gambar dari blok diagram analisis *cepstral*, dengan masukan $s(n)$ merupakan *sequence* dari sinyal wicara yang digunakan. Sinyal *sequence* tersebut kemudian dikalikan dengan *hanning window* dan menghasilkan $x(n)$ yang merupakan sinyal *windowed sequence*. Nilai sinyal $x(n)$ tersebut akan dilakukan proses FFT dan menghasilkan $x(\omega)$ sebagai spektrum dari $x(n)$. *Log magnitude spectrum* didapatkan dengan melakukan proses logaritma dari spektrum $x(\omega)$. Dilakukan invers FFT untuk mentransformasikan kembali sinyal ke domain waktu, namun pada invers FFT *log spectra* ditransformasikan ke domain *frequency* yang sama persis dengan domain waktu. Pemisahan dapat dilakukan dengan melakukan invers FFT kombinasi *log spectra* dari eksitasi dan karakteristik *filter* dari *vocal tract* secara linear [9].

Terdapat pula proses *liftering* yang merupakan operasi yang serupa dengan *filtering* pada domain *frequency*. Terdapat dua macam *liftering* berdasarkan, yang pertama yaitu *high-time liftering* yang memiliki fungsi untuk mendapatkan karakteristik ekstasi dari glotis, sedangkan yang kedua yaitu, *low-time liftering* yang berfungsi untuk mendapatkan karakteristik *vocal tract* yang biasa digunakan untuk *voice recognition* [9]. Pada penelitian ini digunakan *low time liftering*.

2. Pembahasan

Dalam penelitian ini, dilakukan pengujian akurasi performansi sistem dengan menggunakan dua metode yang dibandingkan yaitu analisis *cepstral* dan MFCC. Perbandingan kedua metode tersebut berdasarkan nilai akurasi dalam mengklasifikasi berkas asli, *stego* ataupun *noise*. Terdapat berkas sinyal wicara dengan durasi 10

detik sebanyak 15 berkas dengan ukuran tiap berkasnya 937 Kb dengan *bit rate* 768 Kbps. Dari 15 berkas sinyal wicara asli, diambil 5 sinyal wicara pertama yang disisipkan oleh pesan rahasia dengan metode *Discrete Cosine Transform* (DCT), *Discrete Wavelet Transform* (DWT) dan *Least Significant Bit* (LSB) yang masing-masing metode disisipkan pesan dengan ukuran 359 karakter.

Isi pesan steganografi yang digunakan yaitu: "Steganografi biasa didefinisikan sebagai seni menyembunyikan informasi rahasia dalam pesan lain yang tampaknya tidak berbahaya sehingga mencegah adanya tindakan pendeteksian pesan tersembunyi". Dibandingkan teknik enkripsi lain, steganografi merupakan teknik yang lebih efektif untuk meningkatkan keamanan pesan selama proses transfer data.

Dengan 5 sinyal wicara yang sama, dilakukan penambahan *noise* yang terdiri dari *Additive White Gaussian Noise* (AWGN), *noise gaussian*, dan *noise echo*. Sehingga total berkas sinyal wicara yang digunakan yaitu 45 berkas, 15 berkas asli, 15 berkas *stego* dan 15 berkas *noise*. *Noise gaussian* dibuat dalam bentuk distribusi normal dalam domain waktu dengan nilai domain waktu rata-rata yaitu nol, sedangkan AWGN pada dasarnya sinyal *noise* yang ditambahkan bersifat aditif (*additive*), berwarna putih (*white*) dan terdistribusi normal (*gaussian*). Putih mengacu pada gagasan bahwa sinyal tersebut memiliki kekuatan yang seragam di seluruh pita frekuensi. Sedangkan *noise echo* yaitu dilakukan penambahan gema yang berulang pada berkas sinyal. Tabel 1 menunjukkan *level Signal to Noise Ratio* (SNR) menggunakan tiga jenis *noise* yang sudah disebutkan sebelumnya. Nilai SNR untuk tiap metode *noise* berbeda-beda dari rentang 10 dB hingga 25 dB. Penggunaan level SNR yang berbeda-beda untuk membuktikan kerusakan yang terjadi pada sinyal asli.

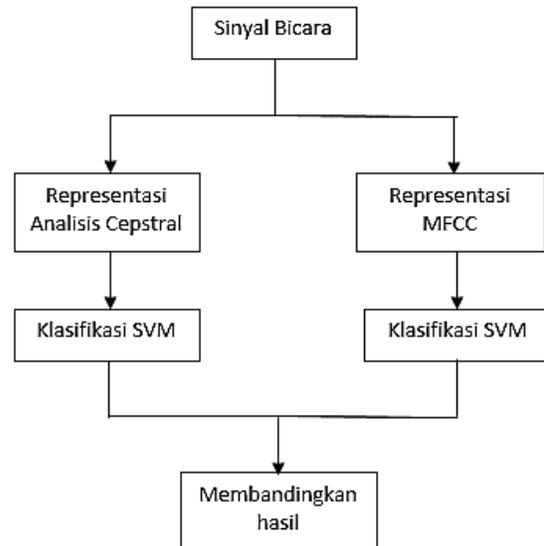
Desain pemodelan sistem pada penelitian ini adalah suatu sistem perancangan pengolahan sinyal audio digital berupa berkas sinyal wicara (*speech*). Konfigurasinya terdiri dari tiga tahap, yaitu: representasi berkas bicara, pengklasifikasian, dan perbandingan hasil, seperti ditunjukkan pada Gambar 3.

Pada tahap pengklasifikasian, dilakukan proses pelatihan terlebih dahulu terhadap data latih kemudian dilakukan pengujian berdasarkan data yang telah dilatih terhadap data uji. Pada tahap pelatihan, hasil dari sinyal wicara yang diolah oleh kedua metode yang dipakai, akan ditarik kesimpulan sementara dari hasil data tersebut, yang dinamakan data latih. Data latih akan dijadikan sebagai *database* latih untuk digunakan pada tahap pengujian.

Tahap pelatihan metode analisis *cepstral* dapat dilihat pada Gambar 4(a). Setelah memilih sinyal wicara latih, dilakukan framing tanpa menggunakan *overlap* pada sinyal tersebut. Proses *framing* tersebut akan

Tabel 1. Nilai SNR pada *Speech Noise*

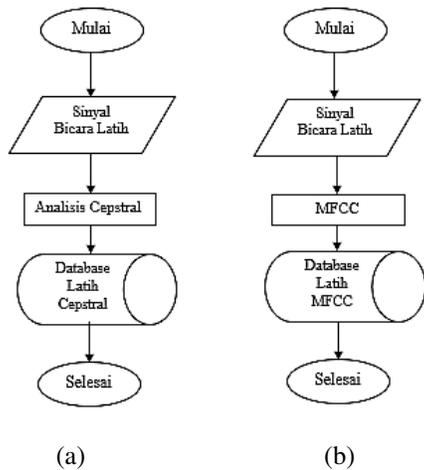
Jenis Berkas	Nilai SNR
AWGN6	14,8707 dB
AWGN7	15,5876 dB
AWGN8	10,7838 dB
AWGN9	12,9772 dB
AWGN10	14,9936 dB
Gaussian6	25,3647 dB
Gaussian7	24,4923 dB
Gaussian8	23,2333 dB
Gaussian9	24,0746 dB
Gaussian10	24,9097 dB
Echo6	14,9570 dB
Echo7	16,6110 dB
Echo8	20,7146 dB
Echo9	19,6350 dB
Echo10	16,9909 dB



Gambar 3. Konfigurasi Umum Sistem

menghasilkan 334 frames dengan panjang satu frame-nya 30 ms dan jumlah *samples* dalam satu buah sinyal wicara sebanyak 480.000 *samples*, sehingga di dalam satu *frame* terdiri dari 1440 *samples*. Sedangkan untuk metode MFCC pada Gambar 4(b), dilakukan *framing* dengan *overlap* dan nilai koreksi M sebesar 100 *samples*, maka didapatkan jumlah *frame* sebanyak 4786 *frames* yang dalam tiap *frame* terdiri dari 1440 *samples* juga.

Setiap frame yang telah dibuat proses *window* kemudian diaplikasikan dengan jumlah *frame*-nya terdiri dari 1440 sampel. Pada metode analisis *cepstral*, jenis *window* yang paling sering digunakan adalah *Hamming Window*, sedangkan untuk metode MFCC jenis *window*



Gambar 4. Tahap Pelatihan Metode Cepstral (a), dan MFCC (b)

digunakan Hanning Window. Perbedaannya terletak pada nilai koefisien window yang digunakan. Pada *Hamming Window* besar koefisiennya 0,54, sedangkan *Hanning Window* adalah 0,5. Nilai dari *window* yang telah dibuat dikalikan ke setiap sampel yang berada di tiap *frame*.

Nilai dari sinyal yang telah diberi *window* akan ditransformasikan dari domain waktu dalam bentuk domain frekuensi, menggunakan fungsi DFT, dengan algoritma FFT. Nilai keluaran dalam domain frekuensi ini memiliki nilai bilangan asli dan bilangan kompleks. Fungsi absolute diberlakukan kepada tiap sampel didalam *frame* untuk menghilangkan nilai bilangan kompleks.

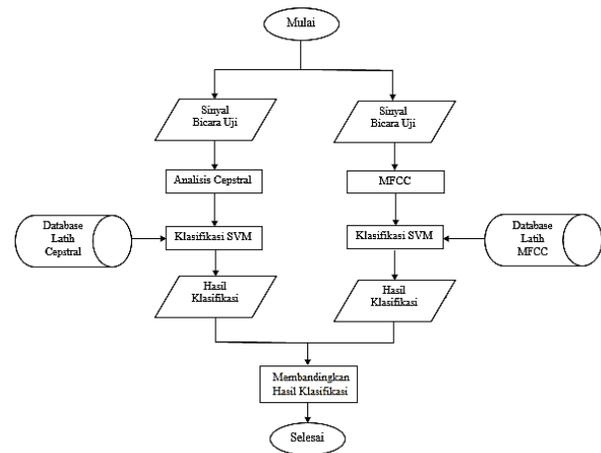
Perbedaan pada kedua metode yang digunakan adalah pada MFCC. Setelah fungsi DFT dilakukan pada sinyal, akan dilakukan proses *Mel-Frequency Wrapping* yang dapat meniru sistem pendengaran manusia, dengan cara mengonversi frekuensi rendah dan frekuensi tinggi yang digunakan ke nilai *mels*. Kemudian dibentuk 20 filter yang diaplikasikan ke setiap *frame*. Sedangkan untuk analisis cepstral, setelah sinyal dilakukan fungsi DFT, setiap sampel di dalam *frame* langsung dilakukan fungsi log, begitu pula dengan sampel yang telah *filter* dalam tiap *frame* untuk MFCC.

Tahapan terakhir, yaitu mentransformasikan kembali sinyal dari bentuk frekuensi ke domain waktu. Pada analisis cepstral digunakan fungsi invers DFT, dengan algoritma invers FFT. Transformasi pada analisis cepstral mentransformasikan sinyal ke domain frekuensi yang sangat mirip dengan domain waktu. Oleh karena itu, *liftering* harus dilakukan, *liftering* identik dengan *filtering* pada domain frekuensi, yang terdiri dari *low-time liftering*, untuk mendapatkan karakteristik *vocal tract* dan *high-time liftering* untuk mendapatkan

karakteristik eksitasi *glottal*. Proses analisis cepstral yang digunakan adalah *low-time liftering*, dengan menerapkan *Low Pass Filter* untuk mengambil nilai pada 15 sampel pertama pada sinyal wicara. Pada MFCC, proses DCT dilakukan untuk mentransformasikan kembali sinyal dari domain frekuensi ke domain waktu sebagai fungsi invers FFT. Nilai ini yang digunakan untuk mendapatkan nilai statistik yang terdiri nilai *skewness*, *kurtosis*, dan *standard deviation* pada tiap *frame*.

Nilai-nilai statistik tersebut disimpan kedalam *database*, untuk selanjutnya dilakukan *Support Vector Machine SVM train* dan dikategorikan berdasarkan kelasnya masing-masing. Kategori yang pertama terdiri dari berkas 'Asli' dan 'Tidak Asli'. Kategori 'Tidak Asli' sendiri terdiri dari berkas *stego* dan *noise*. Untuk kategori yang kedua, berkas untuk kelas *stego* dan *noise* dipisahkan kembali dengan kategori 'Stego' dan 'Noise'.

Database dari masing-masing metode yang didapat dalam tahap pelatihan digunakan dalam proses klasifikasi, menggunakan SVM [10], untuk selanjutnya diketahui apakah sinyal wicara uji tersebut masuk ke dalam kategori berkas *stego*, berkas *non-stego* maupun berkas *noise* seperti ditunjukkan Gambar 5.



Gambar 5. Tahap Pengujian

Proses tahap pengujian sama dengan proses tahap pelatihan. Nilai-nilai statistik yang didapat dari tahap pengujian selanjutnya akan dibandingkan dengan nilai statistik yang telah dilatih dan disimpan pada *database*. *Output* dari sinyal wicara yang diuji selanjutnya akan dikategorikan menjadi berkas 'Asli' atau 'Tidak Asli'. Jika berkas tergolong 'Tidak Asli', maka akan dilakukan klasifikasi untuk menentukan apakah berkas tersebut 'Stego' atau 'Noise'.

Perbandingan akurasi dari kedua metode yang digunakan merupakan proses terakhir untuk menentukan metode mana yang lebih baik. Untuk mengidentifikasi sebuah berkas sinyal wicara apakah memiliki pesan tersembunyi atau bahkan sudah mengalami penambahan

noise.

Performansi sistem dapat dilihat dari hasil parameter akurasi tiap-tiap metode yang telah dibandingkan. Akurasi sendiri merupakan suatu ukuran ketepatan pada sistem dalam mengenali masukan data yang diberikan sehingga akan menghasilkan keluaran yang benar. Nilai akurasi didapatkan dari persamaan (1):

$$Akurasi = \frac{Jumlah\ Data\ Benar}{Jumlah\ Data\ Keseluruhan} \times 100\% \quad (1)$$

Parameter-parameter pengujian yang dilakukan yaitu pengujian nilai akurasi kedua metode terhadap penggunaan *window hamming* dan *window hanning*, pengujian terhadap penggunaan jenis *frame* tanpa *overlap* dan dengan *overlap*, serta pengujian terhadap panjang *frame* yang digunakan yaitu 30 ms, 60 ms dan 90 ms pada kedua metode yang dibandingkan.

2.1 Pengujian Window

Analisis akurasi dilakukan berdasarkan jenis *window* yang diterapkan. *Window* yang digunakan yaitu *Hamming* dan *Hanning*. Perbedaan antara kedua jenis *window* menurut teori, yaitu pada *window hamming* masih memiliki kekurangan, dimana di bagian tepi-nya terjadi diskontinuitas.

Perbandingan nilai akurasi dari tiap jenis *window* didasarkan pada dua klasifikasi, yaitu kategori Asli atau 'Tidak Asli'. Jika "Tidak Asli", maka dikategorikan kembali menjadi 'Stego' atau 'Noise'. Spesifikasi sistem pada pengujian ini menggunakan jenis *frame* tanpa *overlap*, dengan panjang *frame* 30 ms. Tabel 2 dan Tabel 3 berturut-turut menunjukkan hasil penggunaan *window hamming* dan *window hanning*.

Dapat dilihat bahwa akurasi yang lebih baik didapatkan jika sistem menggunakan *window* jenis *hanning*.

2.2 Pengujian Jenis Frame

Setelah diketahui bahwa penggunaan *window hanning* menghasilkan nilai akurasi yang lebih baik, kemudian dilakukan analisis lanjutan, mengenai pengaruh dari jenis *framing* dengan *overlap* atau tanpa *overlap*. Tiap berkas memiliki panjang sebesar 480.000 sampel. Karena panjang yang digunakan adalah 30 ms, maka jumlah sampel dalam satu *frame* adalah sebanyak 1440 sampel.

Untuk sistem yang tanpa *overlap*, didapatkan jumlah *frame* sebanyak 334 *frames*. Hasil akurasi untuk sistem tanpa *overlap* sama dengan hasil pengujian *window hanning* pada pembahasan sebelumnya, seperti ditunjukkan pada Tabel 3. Nilai *overlap* yang baik adalah di atas 50%. Pada percobaan dengan *overlap* ini dilakukan dengan nilai koreksi M sebesar 100 sampel dari tiap *frame* sehingga terbentuk 4786 *frames*. Dengan kata lain *overlap* yang digunakan adalah 93% dari

sample tiap *frame*. Jumlah *frame* yang lebih banyak dari pada tanpa *overlap* dikarenakan terjadinya tumpang tindih antar sample dari *frame* sebelumnya dengan *frame* setelahnya. Fungsi dari dilakukannya *overlap* adalah untuk meminimalisir hilangnya ekstraksi ciri yang dibutuhkan. Hasil akurasi untuk sistem dengan *overlap* dapat dilihat pada Tabel 4. Dapat dilihat bahwa nilai akurasi dengan menggunakan *overlap*, karena dengan adanya *overlap* dapat meminimalisir hilangnya ekstraksi ciri.

Tabel 2. Hamming Window

Jenis Kategori	Metode	Akurasi
'Asli'	Cepstral	62,22%
'Tidak Asli'	MFCC	48,89%
'Stego'	Cepstral	48,89%
'Noise'	MFCC	46,67%

Tabel 3. Hanning Window

Jenis Kategori	Metode	Akurasi
'Asli'	Cepstral	75,56%
'Tidak Asli'	MFCC	75,56%
'Stego'	Cepstral	51,11%
'Noise'	MFCC	62,22%

Tabel 4. Dengan Overlap

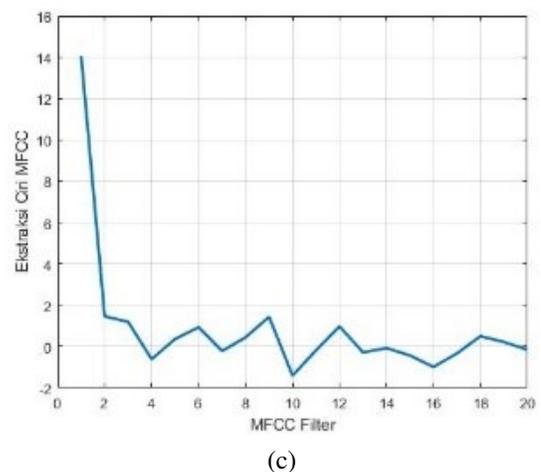
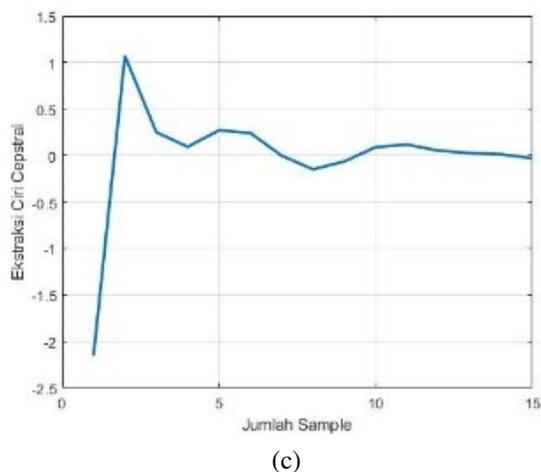
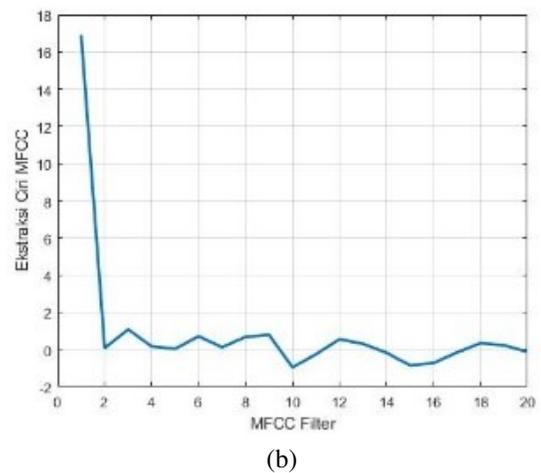
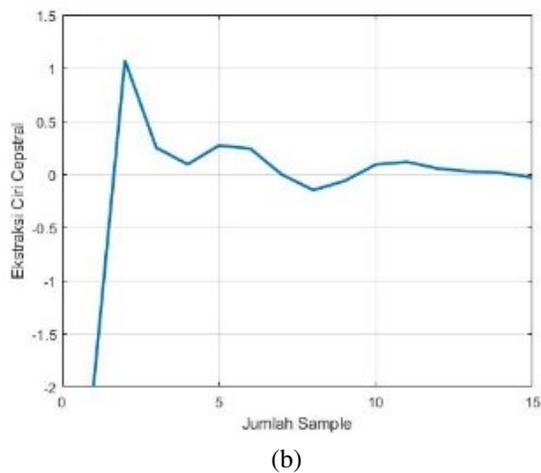
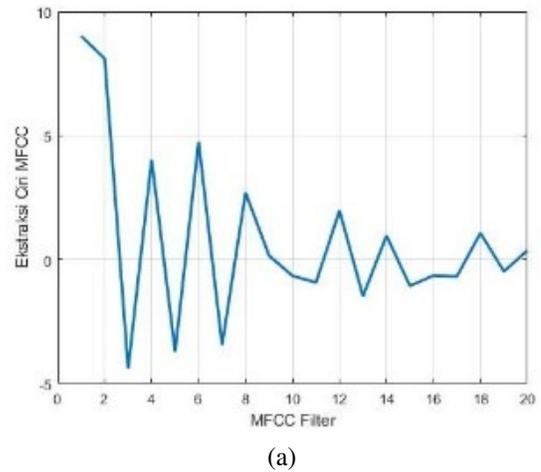
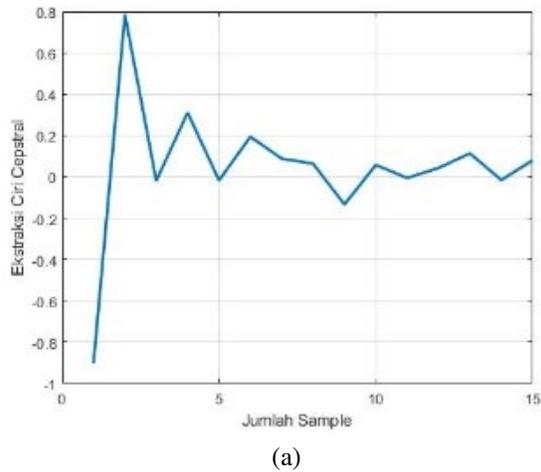
Jenis Kategori	Metode	Akurasi
'Asli'	Cepstral	71,11%
'Tidak Asli'	MFCC	86,67%
'Stego'	Cepstral	46,67%
'Noise'	MFCC	77,78%

Tabel 5. Panjang Frame 60 ms

Jenis Kategori	Metode	Akurasi
'Asli'	Cepstral	68,89%
'Tidak Asli'	MFCC	66,67%
'Stego'	Cepstral	46,67%
'Noise'	MFCC	46,67%

Tabel 6. Panjang Frame 90 ms

Jenis Kategori	Metode	Akurasi
'Asli'	Cepstral	71,11%
'Tidak Asli'	MFCC	75,56%
'Stego'	Cepstral	48,89%
'Noise'	MFCC	51,11%



Gambar 6. Ekstraksi Ciri Analisis Cepstral Berkas Asli (a), Stego DCT (b), dan Noise AWGN (c).

Gambar 7. Ekstraksi Ciri MFCC Berkas Asli (a), Stego DCT (b), dan Noise AWGN (c).

2.3 Pengujian Panjang *Frame*

Dari analisis sebelumnya telah diketahui bahwa penggunaan *overlap* dengan nilai koreksi *M* 100 *samples* menghasilkan akurasi yang lebih baik. Pada pengujian

kali ini akan dilakukan analisis terhadap panjang *frame* yang mempengaruhi nilai akurasi. Dilakukan analisis pada *panjang frame* 30 ms, 60 ms dan 90 ms.

Untuk panjang *frame* 30 ms telah dilakukan pada

pembahasan sebelumnya pada pengujian dengan *overlap* (Tabel 4.), dengan jumlah *frame* sebanyak 4786 *frames*. Jumlah *sample* dalam satu *frame* sebanyak 1440 *samples*. Sedangkan untuk panjang *frame* 60 ms didapatkan jumlah *frame* sebanyak 4772 *frames*. Jumlah *sample* dalam tiap *frame* terdapat 2880 *samples*, dua kali lipat lebih banyak dibandingkan dengan panjang *frame* 30 ms. Dan Untuk panjang *frame* 90 ms didapatkan jumlah *frame* sebanyak 4757 *frames*. Jumlah *sample* dalam tiap *frame* terdapat 4320 *samples*, tiga kali lipat lebih banyak dibandingkan dengan panjang *frame* 30 ms. Hasil nilai akurasi untuk panjang *frame* 60 ms dapat dilihat pada Tabel 5. dan Tabel 6. Untuk panjang *frame* 90 ms.

Dari ketiga hasil akurasi berdasarkan panjang *frame*, nilai akurasi yang paling baik untuk kedua metode yang dibandingkan yaitu ketika digunakannya panjang *frame* 30 ms. Karena kondisi dengan *framing* atau segmentasi pendek (10 ms - 30 ms), bentuk dari nilai *vocal tract* untuk sinyal wicara relatif tetap sama dengan aslinya.

Gambar 6. menunjukkan bentuk pola ekstraksi ciri analisis *cepstral* pada *frame* pertama tiap berkas wicara uji pertama berkas asli (a), *stego* DWT (b) dan *noise* AWGN (c). Sedangkan Gambar 7 menunjukkan pola ekstraksi ciri MFCC pada *frame* pertama tiap berkas wicara uji pertama berkas asli, *stego* DWT dan *noise* AWGN. Berdasarkan bentuk pola yang terlihat dari kedua metode, bentuk pola berkas *stego* dan *noise* terlihat sekilas serupa dibandingkan bentuk pola berkas asli. Oleh karena itu, akurasi proses klasifikasi berkas *stego* dan *noise* dibandingkan berkas asli di setiap parameter uji.

3. Kesimpulan

Pada penelitian ini, telah dilakukan pengujian terhadap berkas sinyal wicara menggunakan dua metode, yaitu analisis *cepstral* dan MFCC dalam beragam parameter, sehingga dapat disimpulkan sebagai berikut:

1. Dari keseluruhan pengujian yang dilakukan dengan metode analisis *cepstral*, spesifikasi sistem yang paling cocok dengan metode tersebut dan menghasilkan akurasi yang paling baik yaitu dengan menggunakan *window hanning*, tanpa *overlap*, dan dengan panjang *frame* 30ms menghasilkan nilai akurasi sebesar 51,11%.
2. Dari keseluruhan pengujian yang dilakukan dengan metode MFCC, spesifikasi sistem yang paling cocok dengan metode tersebut dan menghasilkan akurasi paling baik yaitu dengan menggunakan *window hanning*, dengan *overlap* nilai koreksi M 100 *samples*, dan dengan panjang *frame* 30ms menghasilkan akurasi sebesar 77,78%
3. Perbandingan dari kedua metode yaitu analisis *cepstral* dan MFCC dengan hasil akurasi terbaik

yang dimiliki kedua metode, metode yang dapat mendeteksi berkas sinyal wicara merupakan berkas asli, *stego* ataupun *noise* lebih baik yaitu metode MFCC berdasarkan nilai akurasi terbaik sebesar 77,78% karena pada metode MFCC sudah dapat meniru sistem pendengaran manusia dengan cara mengonversi frekuensi rendah dan frekuensi tinggi yang digunakan pada sinyal wicara ke nilai *mels*, sedangkan untuk analisis *cepstral* nilai akurasi terbaiknya sebesar 51,11%.

Daftar Pustaka

- [1] H. Liu, J. Liu, R. Hu, X. Yan, and S. Wan, "Adaptive audio steganography scheme based on wavelet packet energy," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2017 IEEE 3rd International Conference on.* IEEE, 2017, pp. 26–31.
- [2] V. I. Sinisuka, "Steganalisis audio format wav menggunakan metode discrete wavelet transform dan linear discriminant analysis," 2017.
- [3] S. Whibley, M. Day, P. May, and M. Pennock, "Wav format preservation assessment," Technical Report. British Library. <http://wiki.dpconline.org/images/4/46>, Tech. Rep., 2016.
- [4] M. Z. Aminudin, "Steganalisis file audio dengan metode statistik mfcc," 2015.
- [5] M. Kalamani, D. S. Valarmathy, C. Poonkuzhal, and R. Karthiprakash, "Comparison of cepstral and mel frequency cepstral coefficients for various clean and noisy speech signals," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 1, p. 28972904, Mar 2014.
- [6] A. Wicaksono, S. N. Endah, S. Adhy, and S. Sutikno, "Aplikasi speech recognition bahasa indonesia dengan metode mel-frequency cepstral coefficient dan linear vector quantization untuk pengendalian gerak robot," in *Prosiding Seminar Nasional Ilmu Komputer Undip 2014*.
- [7] M. Likitha, S. R. R. Gupta, K. Hasitha, and A. U. Raju, "Speech based human emotion recognition using mfcc," in *Wireless Communications, Signal Processing and Networking (WiSPNET), 2017 International Conference on.* IEEE, 2017, pp. 2257–2260.
- [8] Y. Rao, C. Vahanesa, C. K. Reddy, and I. M. Panahi, "Two-stage data-driven single channel

- speech enhancement with cepstral analysis pre-processing,” in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on.* IEEE, 2015, pp. 702–706.
- [9] P. Banerjee, B. Chakraborty, and J. Banerjee, “Procedure for cepstral analysis in tracing unique voice segments,” in *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on.* IEEE, 2015, pp. 351–356.
- [10] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, “Penerapan metode klasifikasi support vector machine (svm) pada data akreditasi sekolah dasar (sd) di kabupaten magelang,” *Jurnal Gaussian*, vol. 3, no. 4, pp. 811–820, 2014.