

# MODEL SEGMENT ANTAR PUNCAK YANG BERURUTAN : PENGEMBANGAN MODEL SINUSOIDA UNTUK KOMPRESI SINYAL UCAPAN

Suhartono Tjondronegoro<sup>1</sup>, Florentinus Budi Setiawan<sup>2</sup>

<sup>1,2</sup>Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung

<sup>2</sup>Teknik Elektro, Universitas Katolik Soegijapranata

<sup>1</sup>[shtntjnegoro@stei.itb.ac.id](mailto:shtntjnegoro@stei.itb.ac.id), <sup>2</sup>[fbudisetiawan@yahoo.com](mailto:fbudisetiawan@yahoo.com)

## Abstrak

Pengkodean sinyal ucapan merupakan bagian yang sangat penting dalam sistem komunikasi. Sinyal ucapan dapat dinyatakan sebagai kombinasi dari beberapa sinyal sinusoida. Pada tulisan ini dijelaskan mengenai model sinusoida secara segmental, yang merupakan pengembangan dari model sinusoida. Model sinusoida telah diterapkan pada pengkode sinyal ucapan *Sinusoidal Transform Coding* (STC). Model ini membutuhkan banyak sinyal sinusoida dalam paujaug segmen yang tetap. Model baru yang diusulkan ini adalah model sinusoida yang ditinjau secara segmental antara puncak maksimum menuju puncak minimum berikutnya atau sebaliknya, dalam panjang segmen yang variabel. Berdasarkan model baru, usulan ini, dapat diperoleh parameter puncak dan periode setiap segmen sinyal. Informasi ini yang selanjutnya dikirim ke sisi penerima untuk direkonstruksi menjadi sinyal ucapan estimasi. Berdasarkan hasil percobaan diperoleh nilai SegSNR 24,33 dB.

**Kata kunci :** cosinus, periode, puncak, segmental, sinusoida

## Abstract

Speech coder is one of the most important part of communication systems. Speech signal can be represented as a combination of many sinusoidal signals. On this paper, we describe the segmental sinusoidal model. Sinusoidal model has been applied on Sinusoidal Transform Coding (STC) speech coder. This model needs a large number of sinusoidal components on the fixed segment duration. The new proposed model is a sinusoidal model on the segment between maximum peak and the consecutive minimum peak, and vice versa, on the variable segment duration. Based on the new proposed model, we obtain peak and period parameters for each segment. This information is sent to receiver, then the estimated signal is reconstructed. Based on the experimental results, we obtain the SegSNR of 24.33 dB.

**Keywords :** cosine, peak, period, segmental, sinusoidal

## 1. Pendahuluan

Lebar pita pada sistem telepon digital untuk aplikasi komunikasi ucapan manusia adalah 4 kHz dan dicuplik pada frekuensi 8 kHz. Lebar pita tersebut disesuaikan dengan frekuensi ucapan percakapan manusia pada 300 Hz sampai dengan 3400 Hz, agar sumber ucapan dapat teridentifikasi dengan jelas. Pada tulisan ini, sinyal ucapan percakapan manusia pada uraian selanjutnya hanya disebut sebagai sinyal ucapan. Sinyal ucapan adalah sinyal random yang tidak stasioner dan dianggap *quasi* stasioner pada daerah yang terbatas. Berdasarkan karakteristik tersebut maka sinyal ucapan dapat diolah berdasarkan sifat-sifat statistiknya.

Ucapan manusia terdiri dari bunyi yang bergetar dan tidak bergetar. Bunyi yang bergetar merupakan sinyal yang hampir periodik, yaitu satu segmen sinyal memiliki kemiripan dengan segmen sinyal sebelum atau sesudahnya. Sinyal bergetar relatif lebih mudah dianalisis dan mengandung unsur-unsur *redundant*. Sinyal tidak bergetar lebih sulit dianalisis mengingat karakteristiknya yang mirip sinyal derau [4]. Sinyal bergetar dilambangkan

sebagai huruf vokal dan beberapa huruf konsonan. Sebagian konsonan yang lain melambangkan sinyal tidak bergetar. Konsonan sebagai fonem, membawa informasi yang lebih banyak daripada vocal. Selain itu juga terdapat tipe sinyal ucapan yang bersifat transien [9].

Bagian sinyal bergetar bersifat agak (*quasi*) periodik dengan panjang periode yang sebagai *pitch*. Periode *pitch* menentukan frekuensi dasar dari pembicara sinyal tersebut [10]. Ucapan manusia berupa bunyi yang tidak bergetar memiliki bentuk seperti sinyal acak (*random*). Bentuk sinyal yang menyerupai sinyal random memiliki pita frekuensi yang lebar, tetapi memiliki energi yang lebih kecil daripada energi sinyal bergetar.

Pengkodean ucapan untuk penerapan laju rendah membutuhkan parameter dari ucapan menggunakan analisis dan sintesis sistem. Analisis dapat berupa *loop* terbuka atau *loop* tertutup. Salah satu analisis *loop* tertutup adalah *analysis-by-synthesis*. Parameter diperoleh dengan menghitung nilai *Mean Square Error* antara ucapan asli dengan *factor* pembobotan persepsi dengan ucapan yang dibangkitkan kembali (ucapan sintesis). Kinerja algoritma pengkodean ucapan dinilai berdasarkan

kualitas ucapan, kompleksitas algoritma, penundaan, dan degradasi pada saluran dan derau [3].

Pengkodean ucapan secara garis besar terbagi dalam tiga jenis yaitu pengkodean gelombang, pengkodean parametrik, dan pengkodean hibrid. Selain tiga jenis pengkode tersebut ada jenis yang lain tetapi belum terlalu berkembang yaitu vokoder fonetik. Secara umum fungsi pengkode adalah untuk menganalisis sinyal, menghilangkan faktor-faktor *redundant* dan menyusun kode dari bagian sinyal yang sudah dihilangkan faktor *redundant*nya secara efisien [4].

Penelitian yang cukup banyak dilakukan sampai saat ini adalah pengkode berbasis CELP (*Code Excited Linear Prediction*), model sinusoida dan interpolasi gelombang [7]. Basis CELP masih digunakan untuk dasar menyusun pengkode Ex-CELP (*Extended Code Excited Linear Prediction*) yang merupakan draft standar ITU untuk komunikasi ucapan yang terbaru [14]. Model sinusoida diterapkan pada pengkode sinyal ucapan *Sinusoidal Transform Coding (STC)*. Model interpolasi gelombang diterapkan pada pengkode *Waveform Interpolation Coding (WIC)*. Basis pengkode baru yang diusulkan adalah berbasis sinusoida dengan menerapkan model segmental antar puncak ke puncak, yang belum pernah diteliti sebelumnya.

## 2. Model Sinusoida

Sinyal ucapan dapat dikodekan berdasarkan model sinusoida untuk panjang *frame* antara 15 ms sampai dengan 30 ms. Pada *frame* tersebut, beberapa komponen sinusoida dikirim ke penerima. Jumlah komponen yang dikirimkan berkisar antara 40 sampai dengan 60 sinyal sinusoida untuk dapat membangkitkan sinyal sintesis. Model sinusoida yang diusulkan adalah model sinusoida secara segmental. Sinyal tidak diambil pada segmen yang tetap, melainkan pada segmen dengan panjang yang selalu berubah terhadap waktu. Panjang segmen tergantung dari munculnya puncak maksimum dan puncak minimum sinyal ucapan. Puncak dapat diartikan sebagai nilai maksimum atau nilai minimum sinyal pada selang waktu tertentu, sehingga satu segmen dapat berarti sebagai bagian sinyal dari satu nilai maksimum menuju nilai minimum terdekat berikutnya. Satu segmen dapat berarti juga sebagai bagian sinyal dari satu nilai minimum menuju nilai maksimum terdekat berikutnya. Sinyal diantara puncak maksimum ke puncak minimum berikutnya, atau antara puncak minimum ke puncak maksimum berikutnya dimodelkan sebagai sinyal cosinus selebar setengah periode.

### 2.1 Pemodelan Sinyal dalam Bentuk Sinusoida

Representasi sinyal ucapan dilakukan dengan model pendekatan pembangkitan sinyal ucapan

berupa gelombang eksitasi *glottal* yang telah melalui filter linier yang berubah terhadap waktu. Filter tersebut merupakan model karakteristik resonan dari *vocal tract*. Sinyal ucapan dapat dianggap sebagai eksitasi *glottal* yang terdiri atas dua keadaan, yaitu ucapan bergetar dan ucapan tidak bergetar. Di lain pihak, sinyal ucapan dapat dimodelkan dalam bentuk kombinasi linier komponen sinusoida dengan amplituda, frekuensi dan fasa yang berubah terhadap waktu. Pemodelan dalam bentuk sinusoida telah diusulkan antara lain oleh Almeida et.al dan McAulay dan Quatieri [11]. Pendekatan yang dilakukan oleh Almeida et.al adalah dengan menentukan korelasi antara fasa harmonik di antara potongan-potongan sinyal yang berurutan. Sedangkan McAulay dan Quatieri menggunakan metode *mixed-voicing* dimana fasa dari bagian bergetar diambil dari sampul spektral. Pada keadaan ini fasa bagian tidak bergetar adalah acak. Selubung spektral dinyatakan dalam parameter dengan koefisien prediksi linier. Pemunculan bagian sinyal bergetar yang komponen sinusoida dapat meningkatkan kualitas persepsi [8].

Pada model produksi sinyal ucapan, gelombang sinyal ucapan  $s(n)$  dapat dianggap sebagai keluaran dari filter linier, dengan respons impuls yang berubah terhadap waktu  $h(n)$  yang merupakan model dari karakteristik *vocal tract*. Adapun masukan filter adalah gelombang eksitasi *glottal*  $e(n)$ . Sehingga diperoleh :

$$S(n) = \sum_{l=0}^{\infty} h(n-l).e(l) \quad (1)$$

Sinyal eksitasi *glottal*  $e(n)$  dapat dianggap stasioner untuk selang waktu yang pendek dan dapat dianggap sebagai  $k$  kombinasi linier dari sinyal-sinyal sinusoida dengan amplituda, frekuensi dan fasa yang berbeda. Pada kenyataannya terdapat sebanyak tak terhingga  $k$  komponen sinusoida dalam selang waktu yang terbatas (satu *frame*), sehingga tidak mungkin untuk dilakukan analisis. Pendekatan dilakukan dengan proses transformasi *Fourier short-time* pada satu *frame* untuk mendapatkan puncak-puncak spektra yang terbesar. Sinyal dengan amplituda terbesar, dipakai untuk mewakili keseluruhan sinyal dalam satu *frame*. Semakin banyak  $k$ , maka sinyal asli dapat didekati secara lebih teliti, tetapi menjadi semakin tidak efisien dalam proses analisis. Jika nilai  $k$  dan sinyal sinusoida yang bersesuaian dapat diperoleh, maka dapat dilakukan sintesis untuk mendekati bentuk sinyal asli. Sinyal pendekatan dapat dinyatakan dalam bentuk [11] :

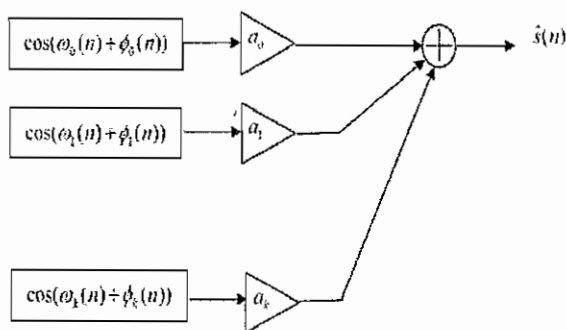
$$\tilde{S}(n) = \sum_{k=0}^K a_k \cos(\omega_k(n)n + \Phi_k(n)) \quad (2)$$

$$1 < K \leq \infty$$

Dimana  $a_k$  menyatakan amplituda sinyal,  $\omega_k(n)$ , menyatakan frekuensi sesaat dan  $\Phi_k(n)$

menyatakan fasa sesaat pada sinyal sinusoida ke- $k$ . Sinyal dalam model ini dianggap dapat dinyatakan dalam  $k$  sinyal sinusoida. Jumlah  $k$  sebanyak tak terhingga dapat ditekan dengan menampilkan bagian yang dianggap mewakili, namun kualitas akan menurun seiring dengan semakin-sedikitnya jumlah bagian sinyal yang mewakili keseluruhan sinyal. Model sinyal secara sinusoida ditampilkan pada Gambar 1.

Sinyal ucapan dapat dipisahkan menurut bentuk periodik dan aperiodik [15]. Sinyal ucapan dapat diuraikan dalam bentuk komponen termodulasi seperti dinyatakan oleh Rao [13]. Selain itu sinyal ucapan dapat juga dimodelkan dalam bentuk modulasi amplituda dan modulasi frekuensi [12]. Pemodelan untuk analisis dan sintesis sistem berdasarkan kombinasi model sinusoida *overlap-add* diusulkan oleh George [5] untuk keperluan sintesis dan perbaikan mutu ucapan. Metode-metode tersebut dipakai sebagai acuan untuk pengkodean sinyal ucapan pada laju lebih kecil atau sama dengan 4 kbps.



Gambar 1. Pemodelan Sinyal Secara Sinusoida

Metode baru yang diusulkan dan dipakai dalam sistem pengkodean usulan adalah model sinusoida untuk kuantisasi berdasarkan puncak ke puncak. Selain itu juga diusulkan metode baru untuk memperbaiki sinyal keluaran dengan beberapa filter *bandpass* yang bersifat adaptif sesuai dengan *formant*.

## 2.2 Pengkodean Sinyal dengan Model Sinusoida

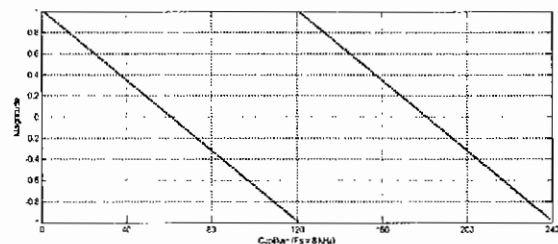
Model sinyal ucapan dalam bentuk sinusoida dapat digunakan untuk mengkodekan sinyal ucapan. Metode pengkodean yang menggunakan model sinusoida adalah *Sinusoidal Transform Coding* (STC). Proses pengkodean transformasi sinusoida bekerja dengan cara mengambil beberapa sinyal sinusoida yang memiliki amplituda yang terbesar. Beberapa sinyal sinusoida dengan amplituda terbesar terlihat pada kawasan frekuensi sebagai bagian spektra yang memiliki magnituda terbesar. Jumlah sinyal sinusoida yang digunakan untuk mewakili sinyal dalam interval waktu tertentu yang disebut sebagai *frame* (15 ms sampai dengan 30 ms) adalah antara 40 sampai 60 sinyal sinusoida.

Proses pengkodean diawali dengan pengambilan sinyal ucapan selebar 15 ms sampai dengan 30 ms dan disimpan dalam *buffer*. Tahap selanjutnya adalah perhitungan spektra sinyal dengan menggunakan transformasi *fourier*. Setelah diperoleh informasi dalam kawasan frekuensi, maka dilakukan pencarian sinyal-sinyal sinusoida yang memiliki amplituda terbesar. Dalam hal ini sinyal sinusoida yang diambil adalah 40 sampai dengan 60 sinyal yang terbesar. Pencarian dilakukan secara bertahap dengan deteksi nilai maksimum lokal secara bertahap, sehingga diperoleh sejumlah sinyal sinusoida yang diinginkan.

Parameter-parameter sinyal berupa nilai-nilai amplituda dan frekuensi dikirimkan ke dekoder untuk dilakukan rekonstruksi, agar diperoleh sinyal estimasi yang diharapkan akan mendekati bentuk sinyal asalnya. Sinyal estimasi dibangkitkan dengan menjumlahkan sinyal-sinyal sinusoida dengan variasi amplituda dan frekuensi. Degradasi kualitas sinyal ucapan terjadi karena adanya pembatasan jumlah parameter sinyal yang dikirim.

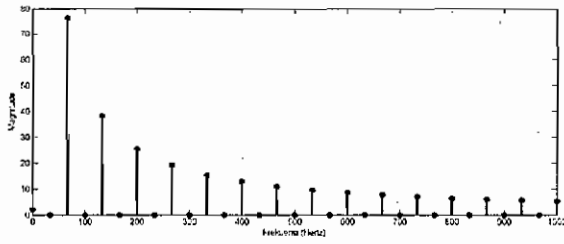
### 2.2.1 Model Sinusoida dalam Segmen Tetap

Model sinusoida telah dipakai untuk pengkodean sinyal ucapan. Sinyal ucapan diambil setiap interval waktu tertentu yang tetap selebar satu *frame*. Pada representasi sinyal di kawasan frekuensi terlihat bahwa terdapat beberapa komponen frekuensi yang cukup menonjol. Komponen-komponen frekuensi yang menonjol inilah yang akan dipakai untuk mewakili sinyal pada interval waktu yang tetap tadi. Semakin banyak komponen frekuensi yang dipakai, maka akan semakin akurat bentuk rekonstruksi sinyalnya. Namun semakin banyak komponen frekuensi yang digunakan, menyebabkan faktor kompresi menjadi turun. Berikut ini adalah contoh sebuah sinyal gigi gergaji yang dengan frekuensi dasar 66,67 Hz, dicuplik pada frekuensi 8 kHz. Dalam satu *frame* terdapat 240 cuplikan sinyal. Gambar 2 menampilkan sinyal tersebut yang berisi dua periode sinyal gigi gergaji.



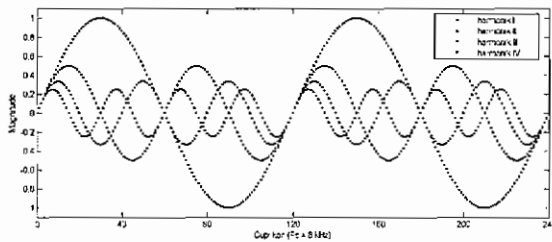
Gambar 2. Sinyal Gigi Gergaji

Representasi sinyal gigi gergaji pada kawasan frekuensi dinyatakan dalam Gambar 3 berikut ini. Berdasarkan informasi pada gambar tersebut, komponen frekuensi dengan daya terbesar adalah 66,67 Hz yaitu frekuensi dasar dari sinyal gigi gergaji.



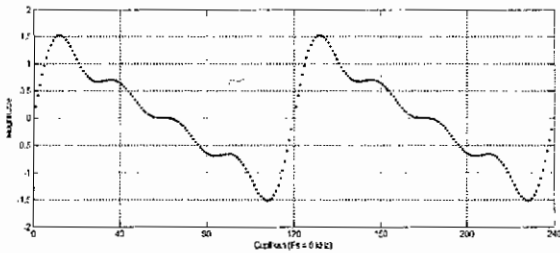
Gambar 3. Representasi Sinyal Gigi Gergaji pada Kawasan Frekuensi

Gambar 4 berikut ini adalah representasi empat buah sinyal pada harmonik pertama, kedua, ketiga dan keempat pada kawasan waktu. Sinyal-sinyal tersebut dijumlahkan dan diharapkan dapat dipakai untuk mewakili sinyal gigi gergaji tersebut di atas.



Gambar 4. Empat Sinyal Harmonik Pertama dari Sinyal Gigi Gergaji

Penjumlahan empat buah sinyal akan menghasilkan sinyal gigi gergaji pendekatan seperti terlihat pada Gambar 5. Sinyal pendekatan berbentuk mirip gigi gergaji yang mengandung ripple.



Gambar 5. Pendekatan Sinyal Gigi Gergaji dengan Empat Harmonik

Tingkat kesalahan semakin kecil jika jumlah sinyal harmonik diperbanyak. Namun dengan semakin meningkatnya jumlah sinyal harmonik, maka untuk transmisi sinyal diperlukan alokasi kanal yang lebih besar untuk menampung sinyal harmonik yang diperlukan. Dengan demikian untuk mendapatkan faktor kompresi yang tinggi, kualitas sinyal harus dikorbankan.

Penggunaan sinyal harmonik seminimal mungkin dapat dilakukan untuk potongan sinyal yang lebih pendek. Sinyal dianalisis berdasarkan segmen di antara puncak, yaitu puncak maksimum menuju puncak minimum berikutnya, atau puncak minimum menuju ke puncak maksimum berikutnya.

## 2.2.2 Model Sinusoida dalam Panjang Segmen Bervariasi

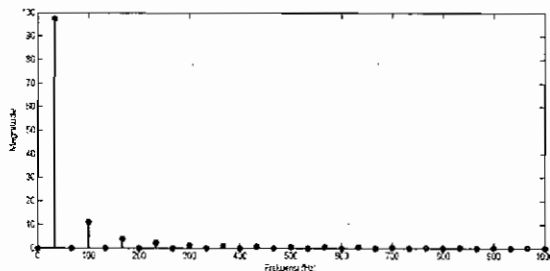
Model sinusoida yang ada sekarang ini menggunakan blok (segmen) dengan panjang yang tetap dan menggunakan sinyal harmonik yang cukup banyak untuk dapat merepresentasikan sinyal dengan kualitas persepsi yang baik. Model baru yang diusulkan adalah merepresentasikan sinyal dalam bentuk dua komponen sinusoida dengan menggunakan panjang blok yang bervariasi. Sinyal ucapan pada dasarnya adalah informasi berupa fluktuasi dari tekanan udara yang masuk ke gendang telinga yang selanjutnya dikirimkan oleh sistem syaraf untuk diproses di otak. Fluktuasi menyebabkan munculnya puncak-puncak maksimum dan puncak-puncak minimum. Bagian sinyal di antara puncak yang berurutan, baik dari puncak maksimum menuju puncak minimum maupun puncak minimum menuju puncak maksimum, mengandung sinyal harmonik yang tak terhingga jumlahnya.

Persepsi pendengaran manusia sangat dipengaruhi oleh tingkat periodisitas sinyal. Apabila posisi puncak sinyal dijaga tetap, maka persepsi pendengaran akan tetap baik. Pada model baru ini, posisi puncak-puncak sinyal tetap dijaga, sedangkan bagian sinyal yang lain direkonstruksi dengan cara menampilkan sinyal harmonik pertama dan sinyal harmonik ke-nol (*offset-DC*). Pemodelan dilakukan dengan mengambil sinyal dengan panjang tertentu yang disebut sebagai *frame*. Pada penelitian ini, panjang *frame* adalah 30 ms atau 240 cuplikan pada frekuensi pencuplikan 8000 Hz. *Frame* tersebut mengandung puncak-puncak maksimum dan minimum yang selalu berubah tergantung sumber ucapan. Sehingga banyaknya puncak yang harus diambil selalu bervariasi.

Parameter yang diperlukan untuk pemodelan adalah puncak-puncak maksimum dan minimum. Bagian sinyal di antara puncak yang berurutan diwakili oleh sinyal harmonik ke-nol dan harmonik pertama. Maka sinyal sinusoida kombinasi harmonik ke-nol dan harmonik pertama selalu berubah untuk setiap segmen diantara puncak yang berurutan. Pada sinyal gigi gergaji seperti penjelasan di atas mengandung dua puncak maksimum dan dua puncak minimum. Puncak-puncak maksimum berada di cuplikan pertama dan cuplikan ke-121. Sedangkan puncak-puncak minimum terdapat pada puncak ke-120 dan cuplikan ke-240. Pada model sinusoida secara segmental usulan baru ini, parameter yang penting adalah puncak-puncak tersebut. Sedangkan bagian yang lain dapat direpresentasikan sebagai sinyal pada harmonik ke-nol dan harmonik pertama.

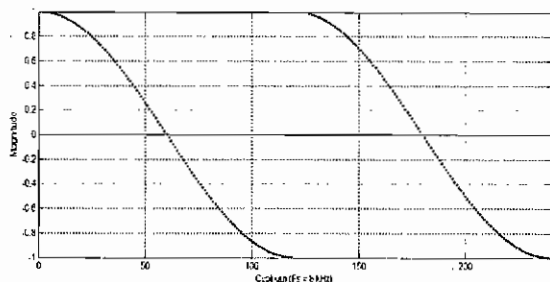
Sinyal gigi gergaji yang diamati pada penjelasan di atas dapat dipandang sebagai dua buah potongan sinyal yang merupakan bagian dari sinyal segitiga. Pada cuplikan pertama sampai dengan cuplikan ke-120, dapat dianggap sebagai setengah periode sinyal segitiga yang memiliki periode

sebesar 240 cuplikan. Gambar 6 berikut ini menunjukkan bahwa pada satu segmen antara puncak maksimum menuju puncak minimum berikutnya mengandung komponen frekuensi yang sangat dominan dibandingkan dengan komponen yang lain. Komponen frekuensi tersebut sama dengan frekuensi harmonik pertama atau frekuensi dasar dari sinyal segitiga. Dengan menggunakan cara yang sama dengan metode sinusoida dalam hal pengambilan komponen frekuensi yang paling menonjol, maka komponen frekuensi pada harmonik pertama saja yang akan dipakai untuk merepresentasikan sinyal.



**Gambar 6. Representasi Sinyal Segitiga pada Kawasan Frekuensi**

Sinyal segitiga selebar setengah periode dapat direkonstruksi menggunakan informasi puncak-puncak maksimum dan puncak-puncak minimum. Pada segmen di antara puncak yang berurutan, sinyal direpresentasikan dengan menggunakan sinyal harmonik ke-nol dan sinyal harmonik pertama. Gambar 7 berikut ini adalah hasil rekonstruksi dengan menggunakan model sinusoida secara segmental yang diusulkan.



**Gambar 7. Rekonstruksi Sinyal Gigi Gergaji yang disusun dari Dua Buah Sinyal Segitiga Setengah Periode dengan Model Sinusoida Secara Segmental**

Ditinjau dari sisi persepsi pendengaran manusia, bentuk sinusoida lebih nyaman didengar oleh telinga daripada bentuk yang mengandung *ripple* yang berlebihan seperti yang terjadi pada model berbasis LPC maupun CELP. Maka metode usulan diharapkan akan relatif lebih nyaman untuk pendengaran. Di lain pihak, dengan menggunakan metode sinusoida secara segmental yang diusulkan, periodisitas sinyal tetap terjaga. Hal ini menjamin akan diperoleh kualitas persepsi yang baik.

### 2.2.3 Model Sinusoida secara Segmental

Model sinusoida dapat dikembangkan lebih lanjut dengan tujuan untuk mendapatkan jumlah parameter yang lebih sedikit agar laju data sinyal ucapan dapat ditekan. Model sinusoida baru yang dikembangkan adalah model sinusoida secara segmental. Model baru ini hanya mengambil satu sinyal sinusoida pada harmonik ke-nol dan harmonik pertama untuk setiap segmen yang dibatasi oleh puncak sinyal pertama dengan puncak sinyal berikutnya yang berlawanan. Puncak dapat diartikan sebagai nilai maksimum atau nilai minimum sinyal pada selang waktu tertentu. Sehingga satu segmen dapat berarti sebagai bagian sinyal dari satu nilai maksimum menuju nilai minimum terdekat berikutnya. Satu segmen dapat berarti juga sebagai bagian sinyal dari satu nilai minimum menuju nilai maksimum terdekat berikutnya.

Perbedaan model sinusoida dengan model sinusoida secara segmental yang diusulkan adalah pada jumlah komponen sinusoida yang dikirim dan panjang segmen sinyal yang diamati. Jumlah komponen sinusoida pada model sinusoida adalah sebanyak 40 sampai 60 komponen sinyal sinusoida. Sedangkan model sinusoida secara segmental hanya menggunakan satu komponen sinusoida saja. Panjang segmen yang diproses untuk model sinusoida adalah 15 ms sampai 30 ms dengan panjang yang tetap. Pada model sinusoida secara segmental, panjang segmen bervariasi tergantung jarak waktu antara puncak maksimum dengan puncak minimum berikutnya, atau jarak waktu antara puncak minimum dengan puncak maksimum berikutnya. Panjang segmen berubah antara satu cuplikan (0,125 ms) sampai dengan 64 cuplikan (8 ms). Berikut ini adalah tabel perbandingan antara model sinusoida dengan model sinusoida secara *segmental*.

**Tabel 1. Perbedaan Model Sinusoida dan Model Sinusoida Secara Segmental**

	Model Sinusoida	Model Sinusoida Segmental
Komponen sinusoida	40 – 60	2
Panjang segmen	Tetap, 15 ms – 30 ms	Bervariasi, 0,125 ms – 8 ms

Pemodelan sinyal dengan pendekatan secara sinusoida secara *segmental*, dapat diproses dengan menggunakan algoritma yang disusun seperti penjelasan berikut ini. Sinyal ucapan yang akan diproses berupa sinyal ucapan yang dicuplik pada 8 kHz dengan kuantisasi 16 bit. Panjang *frame* yang diproses adalah setiap 30 ms, agar dapat mencakup panjang periode *pitch* maksimum yang mungkin pada ucapan manusia selebar 20 ms atau sama dengan 160 cuplikan pada frekuensi pencuplikan 8 kHz. Urutan proses untuk memperoleh nilai puncak dan periode sinyal yang didekati adalah sebagai berikut.

Tahap pertama adalah menentukan satu *frame* sinyal dengan panjang 30 ms (240 cuplikan). Selanjutnya puncak-puncak sinyal ditandai, baik puncak yang positif maupun puncak negatif. Selisih magnituda puncak pertama ke puncak berikutnya ke-*i* disebut sebagai informasi puncak sinyal pendekatan dan disebut sebagai  $p_k(i)$ . Jarak waktu antara puncak pertama dan puncak berikutnya ke-*i* merupakan setengah periode sinyal pendekatan dan disebut sebagai  $p_d(i)$ . Selanjutnya dilakukan proses yang sama untuk puncak-puncak berikutnya.

Dengan menggunakan prosedur di atas, maka diperoleh deretan puncak dan deretan periode untuk satu *frame*. Jangkauan dinamik untuk kedua nilai informasi ini jauh lebih kecil dibandingkan dengan jangkauan dinamik sinyal. Maka jumlah bit yang dialokasikan untuk mengirimkan  $p_d(i)$  dan  $p_k(i)$  lebih kecil dibandingkan dengan mengirimkan sinyal tanpa kompresi. Pada pemodelan ini akan didapatkan deretan periode yang berubah untuk tiap segmen dengan amplituda yang berubah. Apabila amplituda dibuat konstan dengan nilai satu, akan membentuk sinyal termodulasi frekuensi. Nilai periode dan puncak untuk tiap segmen dapat dipakai sebagai informasi yang dikirimkan ke dekoder sehingga dapat menghemat kanal transmisi. Pada bagian dekoder, sinyal dapat direkonstruksi, sehingga mendekati bentuk semula dengan model sinusoida. Adapun  $p_d(i)$  dan  $p_k(i)$  memiliki jangkauan dinamik yang jauh lebih rendah daripada sinyal aslinya, sehingga memungkinkan untuk dilakukan proses kompresi berdasarkan model ini.

### 3. Pendekatan Berdasarkan Puncak ke Puncak

Ucapan dapat didengar oleh telinga karena adanya fluktuasi tekanan udara yang menguat dan melemah. Karakteristik sinyal dengan puncak dan lembah (puncak minimum) dapat digunakan sebagai model untuk pendekatan bentuk sinyal ucapan. Sinyal ucapan dapat dikuantisasi berdasarkan segmen antara puncak maksimum ke minimum dan sebaliknya. Pola puncak ke puncak sangat berperan dalam menentukan tingkat periodisitas sinyal. Sedangkan tingkat periodisitas sinyal sangat menentukan persepsi pendengaran manusia, terutama sinyal bergetar yang merupakan bagian terbesar dari sinyal ucapan [8]. Bagian lain yang juga penting dalam menentukan kualitas persepsi adalah *format* pertama dan kedua. Pada metode transformasi sinusoida, dilakukan pendekatan secara harmonik pada satu *frame* tertentu, dengan mengambil puncak-puncak spektra yang terbesar yang diperoleh dari transformasi *fourier short-time* dan bagian tersebut dianggap mewakili sinyal secara keseluruhan [1].

Metode yang diusulkan adalah proses dengan pendekatan secara kawasan waktu secara segmental. Dalam satu selang waktu  $N$  (*frame*) terdapat  $M$  puncak maksimum dan  $L$  puncak minimum. Puncak maksimum dan puncak minimum pada segmen ke-*i*,

masing-masing dinyatakan sebagai  $p_p(i)$  dan  $p_v(i)$ . Sinyal pada segmen ke-*i* (selang waktu dari  $p_p(i)$  menuju ke  $p_v(i)$ ) mengandung sejumlah besar komponen sinusoida. Demikian pula untuk selang waktu antara  $p_v(i)$  sampai dengan  $p_p(i+1)$  juga mengandung komponen sinusoida yang tak terhingga banyaknya. Jika sinyal pada selang waktu antara  $p_p(i)$  dan  $p_v(i)$  dinyatakan dalam Persamaan (2), maka akan terdapat sejumlah besar komponen sinusoida dan dapat dinyatakan sebagai :

$$S(n) = \sum_{k=0}^K a_k \cos(\omega_k(n)n + \Phi_k(n)) \quad (3)$$

untuk :  $1 \leq K \leq I, n = 0, 1, 2 \dots N-1, I \leq N$

Pendekatan dengan metode sinusoida dilakukan dengan mengambil beberapa komponen sinusoida untuk mewakili keseluruhan komponen sinusoida dalam selang waktu dari  $p_p(i)$  menuju ke  $p_v(i)$ . Sinyal  $s(n)$  sepanjang  $N$  cuplikan dapat didekati dengan  $i$  segmen sinyal sinusoida. Komponen sinusoida pertama dan kedua dipakai untuk mendekati sinyal dari puncak ke puncak berikutnya. Sehingga untuk bagian sinyal pada selang ke-*i*, dapat didekati dengan

$$S_{pv}(i, n) = a_0 + a_1 \cos \omega_i(n) \quad (4)$$

untuk :  $P_p(i) \leq S_{pv}(i, n) \leq P_v(i), 1 \leq i \leq I,$   
 $n = 0, 1, 2 \dots N-1, I \leq N$

Nilai  $a_0$  dan  $a_1$  masing-masing menyatakan koefisien *Fourier* yang ke nol (*offset DC*) dan koefisien pertama yang merupakan amplituda harmonik pertama. Pada bagian sinyal dari  $p_v(i)$  sampai dengan  $p_p(i+1)$  dapat dinyatakan sebagai :

$$S_{vp}(i, n) = a_0 + a_1 \cos(\omega_i(n) + \pi) \quad (5)$$

untuk :  $P_v(i) \leq S_{vp}(i, n) \leq P_p(i+1), 1 \leq i \leq I,$   
 $n = 0, 1, 2 \dots N-1, I \leq N$

Dengan model tersebut, maka satu bagian sinyal dari suatu nilai minimum menuju ke nilai maksimum berikutnya dapat didekati sebagai sinyal cosinus dari  $\pi$  menuju  $2\pi$ . Sebaliknya, bagian sinyal dari suatu nilai puncak maksimum menuju puncak minimum berikutnya dapat didekati sebagai sinyal cosinus dari 0 sampai  $\pi$ . Masing-masing sinyal cosinus dikoreksi dengan penambahan *offset DC* untuk mendapatkan posisi nilai puncak yang sama antara sinyal asli dan sinyal sintesis. Sinyal sintesis secara sinusoida pendekatan diperoleh dari nilai puncak maksimum dan nilai-nilai puncak minimum pada selang pengamatan. Jika  $p_k(i)$  adalah puncak pada segmen ke-*i*, maka nilai koefisien komponen sinusoida berdasarkan bentuk sinyal dari  $p_p(i)$  menuju ke  $p_v(i)$  adalah :

$$a_0 = \frac{p_k(i) + p_k(i+1)}{2} \quad (6)$$

untuk :  $1 \leq i \leq I, I \leq N$

$$a_1 = \frac{p_k(i) - p_k(i+1)}{2} \quad (7)$$

untuk :  $1 \leq i \leq I, I \leq N$

Berdasarkan Persamaan (4) dengan menggunakan nilai koefisien seperti pada Persamaan (6) dan Persamaan (7) diperoleh sinyal pendekatan untuk segmen ke- $i$  antara  $p_p(i)$  menuju  $p_v(i)$  sebagai berikut :

$$S_{pv}(i, n) = \frac{p_k(i) + p_k(i+1)}{2} + \frac{p_k(i) - p_k(i+1)}{2} \cos\left(\frac{2\pi n}{T} + \phi_1(n)\right) \quad (8)$$

untuk :  $1 < i < I, I < N, n = 0, 1, 2 \dots N-I$

Potongan sinyal dari bagian puncak minimum menuju ke puncak maksimum, berlaku Persamaan (5). Sehingga nilai koefisien cosinus berdasarkan bentuk sinyal dari  $p_v(i)$  menuju ke  $p_p(i+1)$  adalah :

$$a_0 = \frac{p_v(i) + p_p(i+1)}{2} \quad (9)$$

$1 \leq i \leq N, n = 0, 1, \dots N-I$

$$a_1 = \frac{p_k(i+1) - p_k(i)}{2} \quad (10)$$

$1 \leq i \leq N, n = 0, 1, \dots N-I$

Dengan menggunakan Persamaan (5) untuk nilai koefisien seperti pada Persamaan (9) dan Persamaan (10) diperoleh sinyal pendekatan untuk segmen antara  $p_v(i)$  menuju  $p_p(i+1)$  sebagai berikut :

$$s_{vp}(i, n) = \sum_{l=1}^k \left[ a_0 - a_1 \cos\left(\frac{(n - n_{v(i)})\pi}{n_{p(i+1)} - n_{v(i)}}\right) \right] \quad (11)$$

untuk :  $1 \leq i \leq N, n = 0, 1, 2 \dots N-I$

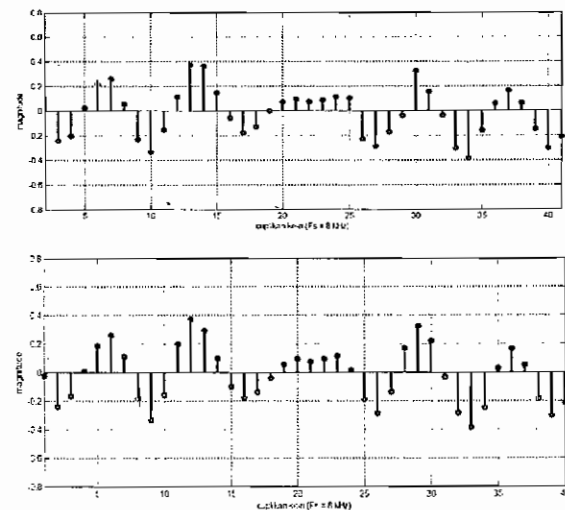
Sinyal pendekatan untuk satu *frame* adalah kombinasi dari kumpulan sinyal dari puncak maksimum menuju ke puncak minimum dan kumpulan sinyal dari puncak minimum menuju ke puncak maksimum. Berdasarkan Persamaan (8) dan Persamaan (11), maka sinyal pendekatan dengan menggunakan harmonik pertama dapat dinyatakan sebagai :

$$S_r(i, n) = \frac{p_k(i) + p_k(i+1)}{2} + \frac{p_k(i) - p_k(i+1)}{2} \cos\left(\frac{\pi(n - n_k(i))}{p_d(i)} + (i+1)\pi\right) \quad (12)$$

Pada potongan sinyal sebesar satu *frame* terdapat  $2k$  segmen sinyal pendekatan sinusoida, masing-masing terdapat sebanyak  $k$  sinyal cosinus dari 0 menuju  $\pi$  dan sebanyak  $k$  sinyal cosinus dari  $\pi$  menuju ke  $2\pi$ .

Informasi periode dan *gain* untuk tiap segmen merupakan informasi yang dikirimkan ke dekoder. Pada bagian dekoder, sinyal direkonstruksi agar mendekati bentuk semula dengan model sinusoida secara segmental. Informasi periode dan puncak

memiliki jangkauan dinamik yang jauh lebih rendah daripada sinyal aslinya, sehingga memungkinkan untuk dilakukan proses kompresi berdasarkan model ini.



**Gambar 8. Sinyal Asal (Atas) dan Rekonstruksi (Bawah) dari Bunyi "aa" Sepanjang empat puluh Cuplikan**

#### 4. Hasil-Hasil Percobaan

Sinyal ucapan dapat dimodelkan dengan pendekatan sinusoida untuk setiap interval atau segmen di antara puncak maksimum dan minimum berikutnya atau sebaliknya. Dengan menggunakan model baru tersebut (sinusoida segmental) akan diperoleh deretan informasi periode dan puncak. Informasi ini dikirimkan ke penerima untuk merekonstruksi sinyal ucapan. Nilai periode dan puncak memiliki jangkauan dinamik yang lebih kecil dibanding dengan sinyal aslinya. Selain itu jumlah informasi periode dan puncak, lebih sedikit dibanding dengan banyaknya cuplikan sinyal asal. Sehingga sangat cocok untuk proses kompresi. Bagian sinyal dari minimum ke maksimum didekati dengan sinyal cosinus dari  $\pi$  menuju  $2\pi$ . Sedangkan bagian sinyal dari nilai maksimum menuju ke nilai minimum didekati dengan sinyal cosinus dari 0 menuju ke  $\pi$ . Percobaan untuk kuantisasi sinyal berdasarkan model sinusoida secara segmental menghasilkan parameter berupa panjang periode dan informasi nilai puncak. Sinyal ucapan yang dicuplik pada frekuensi 8 kHz dengan resolusi konversi analog ke digital 16 bit telah dicoba didalam eksperimen, sinyal-sinyal tersebut dimodelkan dan diambil parameternya.

Kualitas sinyal secara obyektif diukur dengan menggunakan *Segmental signal to noise ratio*. Berdasarkan percobaan terhadap 11 rangkaian kata dalam Bahasa Indonesia yang diucapkan oleh laki-laki dan perempuan diperoleh nilai rata-rata SegSNR sebesar 24,33 dB. Hasil pengukuran MOS

menunjukkan bahwa kualitas persepsi lebih besar dari 4, yang berarti bahwa keluaran yang dihasilkan mirip dengan sinyal asalnya. Rangkaian kata yang sama dengan percobaan di atas, diterapkan pada model sinusoida dengan 66 komponen sinusoida, pada frekuensi pencuplikan 8.000 Hz. Nilai rata-rata SegSNR yang dihasilkan adalah di atas 30 dB. Untuk model sinusoida dengan menggunakan sebanyak 40 komponen sinusoida, menghasilkan SegSNR sebesar 22 dB [2]. Dengan demikian model sinusoida secara segmental menghasilkan sinyal rekonstruksi dengan kualitas yang lebih baik daripada model sinusoida dengan 40 komponen. Di sisi lain, kompleksitas model sinusoida secara segmental jauh lebih rendah dari pada model sinusoida yang telah ada. Model sinusoida membutuhkan parameter amplituda, frekuensi dan fasa. Masing-masing parameter yang disediakan untuk model sinusoida dengan 40 komponen adalah sejumlah 120 parameter.

## 5. Kesimpulan

Sinyal periodik dapat didekomposisikan menjadi sinyal sinusoida dengan jumlah tak terbatas dengan kombinasi amplituda, frekuensi dan fasa. Dalam model sinusoida segmental yang diusulkan ini, dilakukan kuantisasi terhadap nilai puncak-puncak yang diperoleh. Cara ini adalah salah satu metode untuk mengkodekan atau mengkompres sinyal agar didapatkan laju data yang rendah. Parameter sinyal berupa periode dan puncak diperoleh dari nilai-nilai puncak, baik maksimum maupun minimum. Rekonstruksi sinyal dapat dilakukan berdasarkan informasi puncak dan periode. Berdasarkan hasil percobaan awal dengan jumlah responden yang terbatas, untuk nilai SegSNR lebih besar dari 24 dB menghasilkan kualitas MOS lebih besar dari 4. Dengan demikian sinyal rekonstruksi dapat didengar dengan persepsi yang mirip dengan sinyal asalnya dan lebih baik daripada model sinusoida dengan 40 komponen. Kompleksitas model sinusoida secara *segmental* jauh lebih rendah dari pada model sinusoida yang telah ada.

## Daftar Pustaka

- [1] Ahmadi, S., dan A. Spanias, 1998, *A new Model for Sinusoidal Transform Coding of Speech*, IEEE TSAP.
- [2] Alboge, C., dan M.G. Christensen, 2002, *Sinusoidal Speech Modeling and Coding*, Master Thesis, Department of Communication Technology, Institute of Electronics Systems, Aalborg University, Denmark.
- [3] Atal, B.S., V. Cuperman, dan A. Gersho, 1991, *Advances in Speech Coding*, Kluwer Academic Publishers, Massachusetts.
- [4] Furui, S., 1989, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker Incorporation, New York.
- [5] George, E.B dan M.J.T. Smith, 1997, *Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model*, IEEE TSAP, vol 5, hal. 389-406.
- [6] Goldberger, J., D. Burshtein, dan H. Franco, 1999, *Segmental Modeling Using a Continuous Mixture of Nonparametric Models*, *IEEE Transactions on Speech and Audio Processing*, 7, hal. 262-271.
- [7] Gottesman, O., dan A. Gersho, 2001, *Enhanced Waveform Interpolative Coding at Low Bit-Rate*, IEEE Transactions on Speech and Audio Processing, 9, hal. 1-13.
- [8] Painter, T., dan A. Spanias, 2005, *Perceptual Segmentation and Component Selection for Sinusoidal Representation of Audio*, IEEE TSAP, vol.13, hal. 149 – 162.
- [9] Picket, J.M, 1999, *The Acoustics of Speech Communication*, Allyn and Bacon, Boston.
- [10] -----, 1992, *Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction*, ITU-T Recommendation G-728, Geneva.
- [11] Quatieri, T.F., dan R.J. McAulay, 1986, *Speech Transformations Based on a Sinusoidal Representation*, IEEE TASSP, vol. ASSP-34, no. 6.
- [12] Quatieri, T.F., T.E. Hanna, dan G.C. O'Leary, 1997, *AM-FM Separation Using Auditory-Motivated Filters*, IEEE TSAP, vol 5, hal. 465-480.
- [13] Rao, A., dan R. Kumaresan, 2000, *On Decomposing Speech into Modulated Components*, IEEE TSAP vol 8, hal. 240-254.
- [14] Thyssen, J., Y. Gao, dan A. Benyassine, 2001, *A Candidate for The ITU-T 4 kbit/s Speech Coding Standard*, IEEE International Conference on Speech and Signal Processing, hal. 681-684.
- [15] Yegnanarayana, B., C. D'Alessandro, dan V. Darsinos, 1998, *An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components*, IEEE Transactions on Speech and Audio Processing, 6, hal. 1-11.