

# PENKODEAN SUARA PITA LEBAR

Ikhwana Elfitri

Jurusan Teknik Elektro, Fakultas Teknik Unand

## ABSTRAK

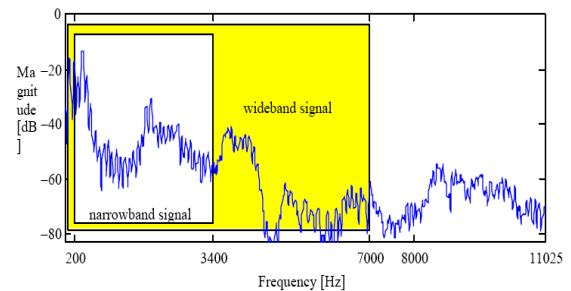
Makalah ini menampilkan studi literatur tentang pengkode suara pita lebar yang ditujukan untuk aplikasi pada sistem komunikasi bergerak generasi ke-tiga (3G). Teknologi 3G telah memberi peluang penggunaan suara pita lebar (frekuensi 50-7000 Hz) untuk meningkatkan kualitas komunikasi suara. Suara pita lebar telah terbukti mampu membuat suara terdengar lebih alami (*naturalness*), memudahkan pendengar membedakan fricative sounds, dan mengurangi tingkat kelelahan dalam berkomunikasi (*listener fatigue*). Perkembangan penelitian tentang metode pengkodean dan metode kuantisasi vektor terhadap LPC parameter pada pengkode suara pita lebar disampaikan beserta algoritma yang digunakan untuk perancangan quantiser vektor.

## 1. PENDAHULUAN

Dalam bidang telekomunikasi, salah satu sistem yang mengalami perkembangan sangat pesat adalah komunikasi bergerak (*mobile communication*) atau yang sering juga disebut dengan komunikasi seluler (*cellular communication*). Generasi pertama (1G) dari sistem ini hanya digunakan untuk komunikasi suara (*voice communication*). Seiring perkembangan teknologi, generasi kedua (2G) sistem komunikasi bergerak dapat digunakan untuk mengirimkan data dalam kapasitas kecil seperti layanan *short message service* (sms). Sekarang, sistem ini memasuki generasi ketiga (3G) yang ditandai dengan kemampuan menyediakan layanan transmisi data berkapasitas besar. Di Indonesia, beberapa operator sistem komunikasi bergerak seperti Telkomsel, Indosat, Excelcomindo dan operator lainnya juga telah menyediakan jaringan yang mendukung teknologi 3G.

## 2. SUARA PITA LEBAR (*WIDEBAND SPEECH*)

Pada sistem komunikasi seluler 3G, ada peluang pesawat telepon menggunakan suara dengan lebar pita frekuensi yang lebih besar [1]. Gambar 1 memperlihatkan bahwa sinyal suara memiliki rentang frekuensi yang cukup besar. Lebar pita frekuensi suara yang biasanya digunakan untuk percakapan telepon adalah pada rentang frekuensi 300 – 3400 Hz, yang biasa dikenal dengan *narrowband speech*. Walaupun suara dengan rentang frekuensi 300-3400 Hz mempunyai kualitas yang dapat diterima (dikenal dengan istilah *toll quality*), ada motivasi yang kuat untuk meningkatkan kualitas suara, terutama dengan adanya teknologi 3G, dengan memperbesar lebar pita frekuensi menjadi 50 – 7000 Hz. Suara pada rentang frekuensi ini biasanya disebut dengan suara pita lebar (*wideband speech*) [2].



Gambar-1. Spektrum frekuensi sinyal suara [3]

JP Adoul dalam [4] mengatakan ada beberapa kekurangan *narrowband speech* yaitu : (1) suara terdengar tidak alami (*naturalness*), (2) orang yang berbicara terasa berada pada jarak yang jauh dan (3) susah membedakan antara huruf-huruf yang terdengar berdesis (*fricative sounds*) seperti /s/ dan /f/. Semua kekurangan ini akan menyebabkan orang yang mendengarkan *narrowband speech* menjadi kelelahan (*listener fatigue*). Karena itulah, muncul motivasi yang kuat untuk menerapkan suara pita lebar pada berbagai aplikasi sistem komunikasi suara [2].

Dengan menurunkan batas bawah rentang frekuensi (*frequency cut-off*) dari 300 ke 50 Hz, suara terdengar lebih alami dan berarti (*fullness*). Sedangkan menaikkan batas atas rentang frekuensi dari 3400 ke 7000 Hz akan memudahkan pendengar untuk membedakan *fricative sounds* [5]. Karl Hellwig dalam [1] menegaskan bahwa lebar pita (*bandwidth*) yang lebih besar akan meningkatkan kualitas suara sehingga dapat mengurangi tekanan (*stress*) selama melakukan percakapan telepon. C. Quinquis dalam [6] membuktikan dengan eksperimen bahwa pengkode suara pita lebar (*wideband speech coder*) menghasilkan suara dengan kualitas yang lebih baik dibandingkan *narrowband speech coder*.

### 3. PENGKODE SUARA PITA LEBAR

Pengkodean suara (*speech coding*) merupakan proses penting dalam sistem komunikasi suara. Fungsinya adalah merepresentasikan sinyal suara dalam suatu deretan bit (*bit stream*) sehingga dapat ditransmisikan secara digital melalui berbagai sistem seperti sistem telepon kabel (*wire telephone*), telepon seluler (*mobile phone*), dan telepon via internet (*voice over internet protocol*). Sampai saat ini, berbagai penelitian telah dilakukan untuk menemukan metode yang efisien dalam merepresentasikan sinyal suara, dan beberapa diantaranya telah direkomendasikan oleh *International Telecommunication Union* (ITU) dan *European Telecommunication Standard Institute* (ETSI) sebagai standar dalam berbagai aplikasi komunikasi suara. Untuk *narrowband speech*, ITU-T mengeluarkan standar pengkode suara seperti MP-MLQ (G.723.1), LD-CELP (G.728), CS-ACELP (G.729). Sedangkan ETSI mengeluarkan standar RPE-LTP (ETSI full-Rate GSM 6.10).

Berbagai penelitian dalam bidang pengkodean suara ini bertujuan untuk mendapatkan metode yang mampu merepresentasikan sinyal suara dalam ukuran yang sekecil mungkin, dengan tetap mempertahankan kualitas dan kejernihan suara. Dengan demikian, sinyal suara dapat ditransmisikan dengan bit rate transmisi serendah mungkin, sehingga dapat menghemat penggunaan bandwidth frekuensi [7]. Sasaran akhir (*ultimate objective*) dari sistem pengkodean suara ini adalah untuk mendapatkan suara dengan kualitas *transparent*, yaitu reproduksi suara yang secara perceptual tidak berbeda dengan suara aslinya [8].

Dengan tersedianya layanan 3G, maka penelitian tentang pengkode suara pita lebar semakin berkembang. Untuk jaringan telepon kabel, ITU-T telah mengeluarkan beberapa standar pengkode suara pita lebar, yang pertama adalah *adaptive differential pulse code modulation* (ADPCM) dengan bit rate 48, 56 dan 64 kbps [9]. Berikutnya adalah *modulated lapped transform* (MLT) yang bekerja pada bit rate 24 dan 32 kbps [10]. Standar terbaru adalah *adaptive multirate wideband* (AMR-WB) dengan bit rate 6,6 – 23,85 kbps [11]. Untuk sistem komunikasi bergerak, Adaptive Multirate Wideband (AMR-WB) juga telah dijadikan standar untuk sistem GSM/WCDMA oleh 3GPP [12]. Sedangkan 3GPP2 mengeluarkan standar pengkode suara pita lebar Variable-Rate Multimode Wideband (VMR-WB) untuk sistem CDMA2000 dengan bit rate rata-rata 5,77 kbps untuk mode 4 [13].

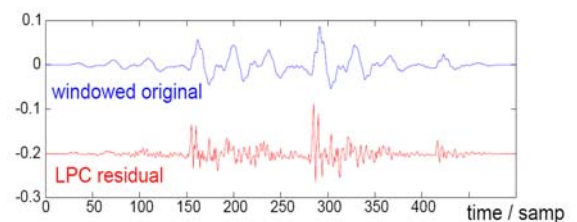
Dari beberapa jenis pengkode suara pita lebar tersebut, yang paling populer adalah AMR-WB, karena pengkode suara ini telah dijadikan standar untuk aplikasi telepon kabel dan telepon seluler, sehingga tidak dibutuhkan *transcoding* dan *tandem operation* untuk koneksi antara kedua sistem tersebut. Disamping itu, AMR-WB juga menghasilkan suara dengan kualitas yang paling

tinggi, yaitu mencapai kualitas *good quality speech* pada bit rate 23,85 kbps [7].

### 4. LPC PADA PENGKODE SUARA PITA LEBAR

Pengkode suara AMR-WB dikembangkan dari pengkode suara *Algebraic Code-Excited Linear Prediction* (ACELP) [14]. Walaupun bekerja dengan multi bitrate, setiap bit rate menggunakan *bitstream* yang sama. *Bitstream* disusun sedemikian rupa sehingga mode dengan bitrate yang lebih rendah dapat ditempatkan (*embedded*) pada mode dengan bitrate yang tinggi. Dengan demikian *receiver/decoder* selalu dapat melakukan rekonstruksi sinyal suara walaupun sebagian dari bit-bit yang berasal dari bitrate yang lebih tinggi tidak dikirimkan [15].

Seperti beberapa pengkode suara lainnya, ACELP menggunakan metode prediksi linier yang dikenal dengan *linear predictive coding* (LPC). LPC digunakan untuk melakukan estimasi selubung spektrum suara (*spectral envelope estimation*) [16]. Salah satu keuntungan menggunakan LPC adalah kemampuannya untuk menurunkan entropy sinyal suara sehingga proses kuantisasi lebih efisien. Hal ini diperoleh dengan cara mendeteksi dan menghilangkan informasi yang dapat diprediksi (*predictable*) dan mengandung perulangan (*redundant*) [8]. Dengan demikian, sinyal suara dapat direpresentasikan dengan jumlah bit yang lebih kecil. Pada kebanyakan pengkode suara, hasil estimasi berupa koefisien LPC dikonversi kedalam parameter *line spectral frequency* (LSF), kemudian dikuantisasi dan ditransmisikan secara digital [2].



Gambar-2. (atas) Sinyal suara masukan untuk analisis LPC (bawah) Sinyal residu keluaran filter LPC [16]

Dewasa ini, kuantisasi koefisien LSF merupakan salah satu topik penelitian utama dalam bidang pengkodean suara. Hampir semua metode terbaru untuk pengkodean suara, menggunakan kuantisasi vektor (*vector quantization*) untuk mengkuantisasi koefisien LSF karena mampu merepresentasikan koefisien LSF dengan jumlah bit yang lebih kecil. Akan tetapi penggunaan kuantisasi vektor secara langsung dibatasi oleh dua masalah yaitu : (1) ukuran memory yang dibutuhkan untuk menyimpan bukukode (*codebook*), dan (2) kompleksitas perhitungan (*computational*

complexity) untuk mendapatkan *codevector* yang optimal [17].

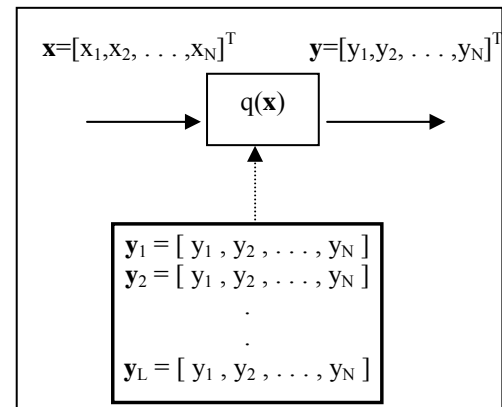
Untuk mengatasi masalah ini maka berbagai penelitian diarahkan untuk mendapatkan metode yang efisien untuk kuantisasi vektor terhadap koefisien LSF pada pengkode suara pita lebar. Salah satu metode kuantisasi vektor yang efektif adalah *split vector quantiser*. Lefebvre dkk [18] menggunakan *split vector quantiser* untuk mengkuantisasi 16 koefisien LSF untuk mendapatkan 48 bit/frame. Chen dkk [19] juga menggunakan *split vector quantiser* untuk mendapatkan 49 bit/frame. Sedangkan Biundo dkk [20] menggunakan *split vector quantiser* untuk mendapatkan 45 bit/frame.

Parameter LSF pada frame yang berurutan, memiliki korelasi yang tinggi [21], karena itu hasil kuantisasi yang lebih baik dapat diperoleh dengan menggunakan *predictive quantiser*. Biundo dkk [20] menggunakan *MA predictive split multistage vector quantiser* (S-MSVQ) untuk mengkuantisasi koefisien LPC menjadi 42 bit/frame. Sementara itu, Guibe dkk [21] menggunakan *safety-net vector quantiser* untuk mendapatkan 38 bit/frame. Beberapa metode terbaru yang digunakan untuk mengkuantisasi parameter LPC pada pengkode suara pita lebar adalah *predictive Trellis-coded quantiser* [22] yang digunakan untuk mendapatkan 34 bit/frame dan *hidden Markov model-based recursive quantiser* [23] yang menghasilkan 40 bit/frame.

## 5. KUANTISASI VEKTOR

Sebuah vektor,  $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$  dimensi-N, akan dikuantisasi secara vektor. Setiap komponen dari  $\mathbf{x}$   $\{x_k, 1 \leq k \leq N\}$  adalah variabel acak yang bernilai riil dengan amplituda kontinyu (*real-valued, continuous-amplitude random variables*). Dengan kuantisasi vektor maka  $\mathbf{x}$  akan dipetakan ke sebuah vektor  $\mathbf{y}$  dimensi-N yang komponennya juga bernilai riil dengan amplituda diskrit. Proses kuantisasi (pemetaan) ini dapat dituliskan sebagai  $\mathbf{y} = q(\mathbf{x})$ , dimana  $\mathbf{y}$  adalah vektor rekonstruksi (*reconstruction vector*) atau vektorkode (*codevector*),  $\mathbf{x}$  adalah vektor masukan (*input vector*), dan  $q$  merupakan operator kuantisasi [24].

Seperti terlihat pada gambar 3, vektorkode  $\mathbf{y}$  merupakan salah satu dari satu set vektor  $\mathbf{Y} = \{\mathbf{y}_i, 1 \leq i \leq L\}$  yang telah tersedia pada memori quantizer, dimana  $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{iN}]^T$ . Set vektor  $\mathbf{Y}$  disebut dengan bukukode (*codebook*). Parameter  $L$  adalah ukuran bukukode dan  $\{\mathbf{y}_i\}$  adalah satu set vektorkode [25].



Gambar-3. Ilustrasi cara kerja vektor quantizer [25]

Salah satu metode dasar yang digunakan untuk disain bukukode adalah algoritma iterasi yang dikenal sebagai algoritma Lloyd (*K-means algorithm* atau *Lloyd algorithm*). Belakangan algoritma ini dikembangkan dan kemudian dikenal dengan algoritma LBG [24].

Secara sederhana, algoritma Lloyd dapat dijelaskan sebagai berikut : (1) Inisialisasi. Set  $m=0$  (iterasi ke- $m$ ). Pilih satu set vektorkode  $\mathbf{y}_i(0), 1 \leq i \leq L$ . (Sebuah bukukode awal). (2) Klasifikasi. Kelompokkan satu set vektor training  $\{\mathbf{x}(n), 1 \leq n \leq M\}$  kedalam  $L$  sel dengan menggunakan aturan nearest neighbor :  $\mathbf{x} \in C_i(m)$ , jika  $d[\mathbf{x}, \mathbf{y}_i(m)] \leq d[\mathbf{x}, \mathbf{y}_j(m)]$ , untuk semua  $j \neq i$ . (3) Perbaharui vektorkode. Ganti  $m$  menjadi  $m+1$ . Hitung vektorkode yang baru untuk setiap sel menggunakan prinsip centroid.  $\mathbf{y}_i(m) = \text{centroid}[C_i(m)]$ ,  $1 \leq i \leq L$ . (4) Tes untuk terminasi iterasi (distorsi  $D(m)$ ) pada iterasi ke- $m$  relatif terhadap  $D(m-1)$  lebih kecil dari nilai treshold tertentu).

Dengan algoritma LBG, bukukode awal dibentuk dengan teknik *splitting*. Dilakukan proses *splitting* terhadap nilai centroid  $y_1^{(1)}$  dari keseluruhan vektor training, sehingga diperoleh dua vektorkode  $y_1^{(1)+\epsilon}$  dan  $y_1^{(1)-\epsilon}$ . Kedua vektor kode ini dijadikan sebagai nilai awal untuk menjalankan algoritma Lloyd, sehingga diperoleh 1 bit vector quantiser,  $y_1^{(2)}$  dan  $y_2^{(2)}$ . Selanjutnya dilakukan proses *splitting* terhadap kedua vektorkode tersebut sehingga diperoleh 2 bit vector quantiser,  $y_1^{(3)}, y_2^{(3)}, y_3^{(3)}$  dan  $y_4^{(3)}$ . Proses ini diteruskan sehingga diperoleh quantiser yang sesuai [26].

## 6. KESIMPULAN

Pada makalah ini telah ditampilkan perkembangan penelitian terbaru (*state of the art*) dan peluang riset pada dari pengkode suara pita lebar yang mencakup tentang metode pengkodean dan metode kuantisasi parameter LPC.

## DAFTAR PUSTAKA

1. Hellwig, K., 2007, Wideband Speech Telephony : from 1984 to 2007, ETSI Workshop on Speech and Noise in Wideband Speech Coding, Sophia Antipolis, France.
2. So, S., Paliwal, K.K., 2007, A comparative study of LPC parameter representations and quantisation schemes for wideband speech coding", Digital Signal Processing, Vol. 17, No. 1, pp. 114-137.
3. VoiceAge, Wideband Speech Coding Standards and Applications.
4. Adoul, J.-P., Lefebvre, R., 1995, Wideband Speech Coding, in : Kleijn, W.B., Paliwal, K.K. (Eds.), Speech Coding and Synthesis, Elsevier, Amsterdam, pp. 289-309.
5. Paulus, J.P., Schnitzler, J., 1996, 16 kbits/s wideband speech coding based on unequal subbands, in : IEEE International Conference on Acoustics, Speech, Signal Processing, pp. 255-258.
6. Quinquis, C., 2007, Quality comparison of wideband coders including tandeming and transcoding, ETSI Workshop on Speech and Noise in Wideband Speech Coding, Sophia Antipolis, France.
7. Gibson, J.D., 2005, Speech Coding Methods, Standards, and Applications, IEEE Circuits and Systems Magazine, Vol. 5, No. 4, pp. 3049.
8. Nurminen J., 2002, Pitch-cycle Waveform Quantization in A 4 kbps WI Speech Coder, Master Thesis, Tampere University of Technology, Finlandia.
9. ITU-T Recommendation G.722, 1988, 7 kHz audio coding within 64 kbits/s.
10. ITU-T Recommendation G.722.1, 1995, Coding at 24 and 32 kbits/s for hands-free operation in systems with low frame loss.
11. ITU-T Recommendation G.722.2, 2002, Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB).
12. 3GPP TS 26.190, 2000, adaptive multirate wideband speech transcoding, 3GPP Technical Specification.
13. Jelinek, M. et al, 2004, On the architecture of the cdma2000 variable-rate multimode wideband (VMR-WB) speech coding standard, Proceeding International Conference on Acoustics, Speech, Signal Processing, pp. I.281-I.284.
14. Besette, B. et al, 2002, The adaptive multirate wideband speech coder (AMR-WB), IEEE Transaction on Speech and Audio Processing, vol. 10, pp. 620-636.
15. Byun, K.J. et al, 2005, An Embedded ACELP Speech Coding Based on The AMR-WB Codec, ETRI Journal, volume 27, Number 2, pp. 231-234.
16. Ellis D., 2006, Speech and Audio Processing & Recognition : speech modeling, Course Notes, Columbia University.
17. Kang S., Song C., Sung H., 2001, A Fast Search Algorithm for the Predictive Split VQ of LPC Parameters, IEEE Communications Letters, Vol 5 No 5.
18. Lefebvre R., Salami R., Laflamme C., Adoul J.-P., 1994, High quality coding of wideband audio signals using transform coded excitation (TCX), Proc. IEEE International Conference on Acoustics, Speech, Signal Processing, pp. 193-196.
19. Chen J.H., Wang D., 1996, Transform predictive coding of wideband speech signals, Proc. IEEE International Conference on Acoustics, Speech, Signal Processing, pp. 275-278.
20. Biundo G. dkk, 2002, Design techniques for spectral quantization in wideband speech coding, Proc. of 3rd COST 276 workshop on Information and Knowledge Management for Integrated Media Communication, Budapest, pp. 114-119.
21. Guibe G. How H.T., Hanzo L., 2001, Speech spectral quantisers for wideband speech coding, European Trans. Telecommunication 12 (6), pp. 535-545.
22. Shin Y. dkk, 2004, Low-complexity predictive trellis coded quantization of wideband speech LSF parameters, in : IEEE International Conference on Acoustics, Speech, Signal Processing, pp. 145-148.
23. Duni E. R. et al, 2004, Improved quantization structures using generalised HMM modelling with application to wideband speech coding, Proc. IEEE International Conference on Acoustics, Speech, Signal Processing, pp. 161-164.
24. Gersho A., Gray R.M., 1992, Vector Quantization and Signal Compression, Kluwer Academic Publisher.
25. Elfritri I., 2002, Pengkode Suara dengan Kuantisasi Vektor, Thesis Magister, Institut Teknologi Bandung.
26. So, S., Paliwal, K.K., 2007, Efficient product code vector quantization using the swiched split vector quantiser, Digital Signal Processing, Vol. 17, No. 1, pp. 138-171.