論文 / 著書情報
Article / Book Information

Title	Relationship between Consecutive Frames in Generalized Harmonics Analysis for Predictive Coding
Author	Hisayori Noda, AKINORI NISHIHARA
Journal/Book name	, , ,
発行日 / Issue date	2009, 5
権利情報 / Copyright	(c)2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Relationship between Consecutive Frames in Generalized Harmonics Analysis for Predictive Coding

Hisayori Noda
Communications and Integrated Systems
Tokyo Institute of Technology
Email: hisayori@nh.cradle.titech.ac.jp

Akinori Nishihara
The Center for Research And
DeveLopment of Educational technology
Tokyo Institute of Technology
Email: aki@cradle.titech.ac.jp

***Abstract*— Relationship between consecutive frame in the generalized harmonics analysis and possibility of predictive coding is described. In the GHA, each frame of a signal is shown as sum of sinusoids. The parameters which show the same sinusoid are often extracted from two consecutive frames. We propose a method to extract the parameters which show the same sinusoid using cross correlation function and greedily selection algorithm. This algorithm is applied to several audio signals and it is confirmed that sinusoids are extracted across multiple consecutive frames. Possibility to apply the proposed method to predictive coding is also shown.**

I. Introduction

Digital audio compression methods are used widely now. They are often used in limited bandwidth like internet streaming and cellular phones. Higher audio compression is required in those information carriers.

Generalized harmonic analysis (GHA) [1] is a concept of signal analysis introduced by N. Wiener in 1930, in which a target signal is expressed as sum of sinusoids. Each sinusoid have three parameters; frequency, amplitude, and phase. Unlike short-time Fourier Transform, the frequency is not restricted to multiples of inverse of the frame size. So the frequency resolution is very high by its nature, and the time resolution is also high because the frame size can be made short without affecting the frequency resolution.

Hirata [2] and Nakazawa [3] applied GHA to the audio coding. Those methods encode each frame independently. On the other hand, the predictive coding using the similarity of two consecutive frames is actively researched in recent years. The compression ratio of audio coding will be increased when predictive coding are applied to GHA audio coding.

The outputs of GHA amplitude, frequency and phase parameters of sinusoids which construct each frame often have the same or similar values in two consecutive frames. These parameters can be used as the similarity between frames for the predictive encoding.

We propose a method to extract the same sinusoid from two consecutive frames from the outputs of GHA. At first, the degree of similarity of each sinusoid pair is calculated and the similarity matrix is generated. The cross correlation function is used for the degree of similarity of two sinusoids. Next, pairs are selected with higher similarity to lower greedily. At last, this algorithm is applied to multiple consecutive frames and long sinusoids are to be extracted.

In the experiment, we applied the proposed method to the target signal and visualized the sinusoids which are detected as the same sinusoid.

We also experimented about predictive coding. In this paper, only the summary of the predictive coding is described. We will describe the detail next time.

II. Generalized Harmonics Analysis

A target signal is divided into frames having size N. The signal in a frame is approximated by the sum of sinusoids as

$$x_0(n) \quad \cong \quad \sum_{k=1}^{K} A_k \sin(\omega_k n + \phi_k), \tag{1}$$

where A_k, ω_k and ϕ_k are the amplitude, the angular frequency and the phase of the k-th sinusoid, respectively, and K is the number of sinusoids to be extracted. We can easily synthesize the signal using the sinusoid parameters. We estimate these parameters so as to minimize the difference between the target signal and the synthesized signal.

Most algorithms for GHA extract sinusoids from a target frame one by one. When a sinusoid is extracted, the parameters of the sinusoids are estimated to minimize the energy of the residue signal. Popular algorithms for the GHA are ABS method [4], Ushiyama's method [5] and Hirata's method [2]. Above algorithms commonly extract sinusoids one by one, but are different in the way of estimating the parameters of a sinusoid.

We proposed an improved algorithm for GHA in a previous research. It is described below.

A. Frequency estimation

In the first step, we estimate the frequency (angular frequency) of a sinusoid to be extracted. In this step, we use discrete time fourier transform (DTFT) of the frame defined by

$$X_k(\omega) = \sum_{n=0}^{N-1} x_k e^{-i\omega n}. \tag{2}$$

978-1-4244-3828-0/09/$25.00 ©2009 IEEE

We search ω which maximizes $|X_k(\omega)|$ defined by

$$|X_k(\omega)| = \sqrt{X_{k_r}^2(\omega) + X_{k_i}^2(\omega)} \quad (3)$$

$$X_{k_r}(\omega) = \sum_{n=0}^{N-1} x_k(n)\cos(\omega n) \quad (4)$$

$$X_{k_i}(\omega) = \sum_{n=0}^{N-1} x_k(n)\sin(\omega n), \quad (5)$$

where $X_{k_r}(\omega)$ and $X_{k_i}(\omega)$ are the real and imaginary part of (2).

Instead of DTFT, we apply FFT to a target signal to get sampled frequencies. Among FFT spectra a peak is detected and call it $\omega(0)$. The truly optimal ω is considered to be around $\omega(0)$. So the range from $\omega(0) - 2\pi/N$ to $\omega(0) + 2\pi/N$ is searched to find the optimal ω, where N is the frame size.

We assume $|X_k(\omega)|$ is convex upward in this range so that

$$\frac{\partial |X_k(\omega)|}{\partial \omega} = 0 \quad (6)$$

holds at the maximum ω.

The derivative is calculated as

$$\begin{aligned} \frac{\partial}{\partial \omega}|X_k(\omega)| &= \frac{\partial}{\partial \omega}\sqrt{X_{k_r}^2(\omega) + X_{k_i}^2(\omega)} \\ &= \frac{X_{k_r}(\omega)\frac{\partial}{\partial \omega}X_{k_r}(\omega) + X_{k_i}(\omega)\frac{\partial}{\partial \omega}X_{k_i}(\omega)}{\sqrt{|X_k(\omega)|}} \\ &= 0, \end{aligned} \quad (7)$$

where $\frac{\partial}{\partial \omega}X_{k_r}(\omega)$ and $\frac{\partial}{\partial \omega}X_{k_i}(\omega)$ are expressed as

$$\frac{\partial}{\partial \omega}X_{k_r}(\omega) = -\sum_{n=0}^{N-1} nx(n)\sin(\omega n) \quad (8)$$

$$\frac{\partial}{\partial \omega}X_{k_i}(\omega) = \sum_{n=0}^{N-1} nx(n)\cos(\omega n). \quad (9)$$

$\frac{\partial}{\partial \omega}|X_k(\omega)|$ is a monotonically decreasing function in this range, and ω can be found by binary search. We can narrow the search range by $1/2$ in one iteration. So calculation accuracy of ω is 2^{-M} where M is the number of iteration. In other words, we can estimate ω in time complexity $O(-N\log\varepsilon)$ where ε is tolerance of frequency.

When we search ω by binary search, the numerator of (7) is unimportant, because only the sign of (7) is considered in the binary search.

The searched ω is the k-th frequency ω_k.

This algorithm may not work well when there are more than two local maxima of $|X_k(\omega)|$ in the range from $\omega(0) - 2\pi/N$ to $\omega(0) + 2\pi/N$. Otherwise it works well even with some noises, Gaussian noise, for example, is mixed in the target signal. In that case, not only sinusoids which construct the target signal but also sinusoids which construct the noises are extracted.

B. Phase and Amplitude Estimation

After estimation of the frequency, we can estimate the phase of the sinusoid simply by

$$\phi_k = \arctan\left(\frac{X_{k_i}(\omega_k)}{X_{k_r}(\omega_k)}\right). \quad (10)$$

To estimate the amplitude of the sinusoid, we use the square sum of the residue. We estimate the amplitude to minimize E_k, expressed by

$$E_k = \sum_{n=0}^{N-1}\{e_k(n)\}^2 \quad (11)$$

$$e_k(n) = x_{k-1}(n) - A_k\sin(\omega_k n + \phi_k). \quad (12)$$

To minimize it, we calculate $\frac{\partial E_k}{\partial A_k} = 0$. It is derived as

$$\begin{aligned} \frac{\partial E_k}{\partial A_k} &= \frac{\partial}{\partial A_k}\sum_{n=0}^{N-1}\{x_{k-1}(n) - A_k\sin(\omega_k n + \phi_k)\}^2 \\ &= \sum_{n=0}^{N-1}\{2(x_{k-1}(n) - A_k\sin(\omega_k n + \phi_k)) \\ &\quad \sin(\omega_k n + \phi_k)\} \\ &= 0 \end{aligned} \quad (13)$$

$$A_k = \frac{\sum_{n=0}^{N-1} x_{k-1}(n)\sin(\omega_k n + \phi_k)}{\sum_{n=0}^{N-1}\sin^2(\omega_k n + \phi_k)}. \quad (14)$$

After calculating parameters of the sinusoid, we subtract it from the target signal. We repeat these steps until enough number of sinusoids are extracted.

C. Time Complexity

Time complexities for FFT, frequency estimation, phase estimation, amplitude estimation and subtraction are $O(N\log N)$, $O(-N\log\varepsilon)$, $O(1)$, $O(N)$ and $O(N)$, respectively. We repeat sinusoid extraction for K times. So the time complexity of the proposed method is

$$\begin{aligned} & O(K(N\log N - N\log\varepsilon + 1 + N + N)) \\ = & \quad O(NK\log\tfrac{N}{\varepsilon}). \end{aligned} \quad (15)$$

III. Extracting the same sinusoid

In GHA, the target signal is divided into frames with frame size N and sinusoids are extracted from each frame. If the target signal is stationary, the same sinusoid, whose amplitude and frequency are similar and the waveforms are consecutive, is to be extracted from two consecutive frames.

In this section, we propose a method to extract the same sinusoid from two consecutive frames. We also propose a method to extract long sinusoids from multiple consecutive frames.

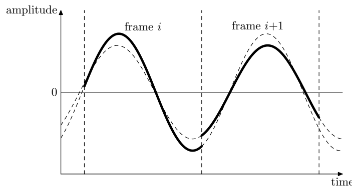

Fig. 1. Degree of similarity of two sinusoids

A. Consecutiveness of sinusoids

We define the degree of similarity of two sinusoids in two consecutive frames. The definition is that the amplitude and the frequency of the two sinusoids are similar and the waveforms are consecutive.

The degree of similarity is defined by the function $S(s_{i,k}(n), s_{i+1,l}(n))$ which use the cross correlation function as

$$s_{i,k}(n) = A_{i,k}\sin(\omega_{i,k}n + \phi_{i,k}), \quad (16)$$

$$F_{ccf} = \frac{1}{A_{i,k}A_{i+1,l}}\sum_{n=0}^{2N-1}(s_{i,k}(n)s_{i+1,l}(n-N)) \quad (17)$$

$$F_a = min(\frac{A_{i,k}}{A_{i+1,l}}, \frac{A_{i+1,l}}{A_{i,k}}), \quad (18)$$

$$F_f = min(\frac{\omega_{i,k}}{\omega_{i+1,l}}, \frac{\omega_{i+1,l}}{\omega_{i,k}}), \quad (19)$$

$$S = F_{ccf}F_aF_f, \quad (20)$$

where $s_{i,k}$ is the k-th extracted sinusoid, $A_{i,k}$C$\omega_{i,k}$ and $\phi_{i,k}$ are the amplitude, the frequency and the phase of $s_{i,k}$ respectively, F_{ccf} is correlation function of two sinusoids whose waveforms are extended to each other's frame, F_a and F_f are the similarity of the amplitudes and the frequencies, respectively. The value of the expressions will increase when the amplitudes and the frequencies of the two sinusoids are similar and the waveform is consecutive, and decrease otherwise.

B. Matching of sinusoid pair

A similarity matrix is generated to apply the above algorithm to each pair of sinusoids in two consecutive frames. The higher values in the matrix are selected without duplication in each row and column to match sinusoid pairs. Negative values and values whose frequency change difference is big are not selected.

C. Consecutiveness from multiple consecutive frames

Applying the above algorithm to multiple consecutive frames and connecting sinusoid pairs which are selected as the same sinusoid following time frames, long sinusoids are extracted.

IV. Predictive Coding

In long sinusoids extracted from multiple consecutive frames, amplitude changes, frequency changes and phase differences in the boundaries between frames will be small. Predictive coding can be applied using these characteristics.

We propose a predictive coding method. At first, the first frame (I frame) of each extracted long sinusoid is processed. The amplitudes and the frequencies are logarithmic quantized and the phases are linear quantized.

After that, the following frames (P frame) are processed. About the amplitudes and the frequencies, the ratio of the current frame to the previous frame are logarithmic quantized. The differences of the current phases and phases predicted from the previous phases and the current frequencies are linear quantized.

At last, the quantized amplitudes, frequencies and phases of I/P frames and the starting frame indexes of the sinusoids are lined up and coded. When the starting frame indexes are lined up, the sinusoids are sorted by the starting frame indexes and the difference between the current and the neighbor are processed.

V. Experiment

A. Method

In the experiment of long sinusoids extraction, we applied the proposed method to the audio signals and checked that long consecutive sinusoids are extracted.

An algorithm proposed by our past research is used for GHA. We implemented the GHA algorithm with CUDA [6], which is programing environment for GPGPU [7] provided by NVIDIA Corp, The frame size is 512 and the number of sinusoids to be extracted is 16.

We analyzed a audio signal of a popular music whose format is 44100HzC16bitC2ch and PCM. Only the left channel of the audio signals is used for analysis.

In the experiment of audio compression, we compressed a speech audio whose format is 44100Hz, 16bit, 1ch and PCM. A band pass filter is applied to the audio to limit the band. The passband is from 200Hz to 2kHz. Eight sinusoids are extracted by GHA. Range coder is used for coding. The data rate is variable and the compression parameters are hard-coded.

B. Results

The results are shown in Figures 2 - 4.

The green dashed lines in Figure 2 are the results of GHA. Each dashed line shows the sinusoid extracted from each frame and contains the parameters of amplitude, frequency and phase. The red solid lines in Figure 2 show the matching results by the proposed method. Each solid line shows long consecutive sinusoids.

TABLE I
Compression Result

Source	PSNR(db)	Data Rate (kbps)
Band Passed Speech Signal	23.38	705.6
Compressed and Decoded Signal	24.17	4.4
G.729	25.12	8.0

In Figure 2, some red lines are intermissive. These sinusoids are not matched to other sinusoids because no positive value left in the greedily matching.

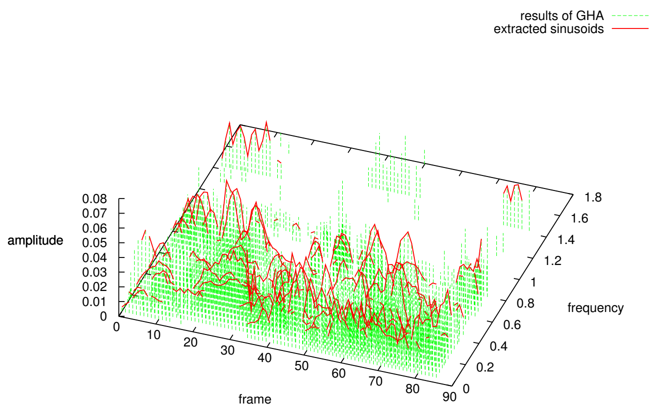

Fig. 2. Result

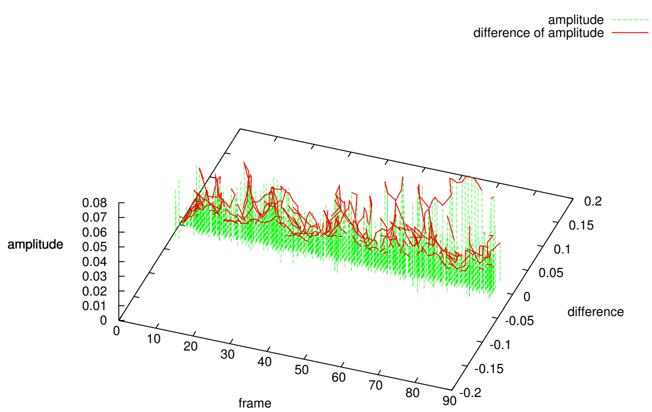

Fig. 3. Difference of amplitudes

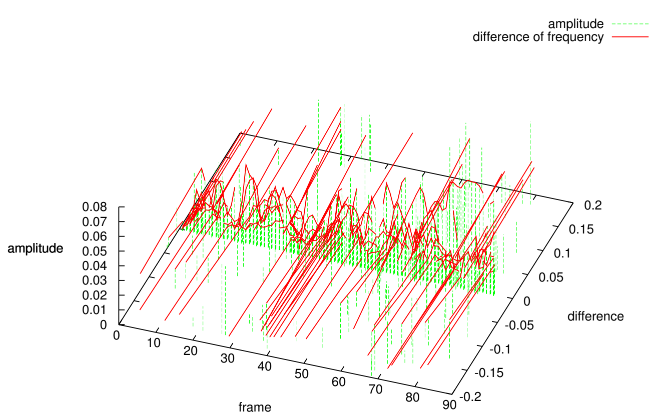

Fig. 4. Difference of frequencies

Figures 3 and 4 show the amplitude differences and the frequency differences of the extracted sinusoids between consecutive frames. The red lines show the differences between frames and the green dashed lines show the amplitudes of the extracted sinusoids.

In Figure 3, the differences of the amplitude gather near 0. This shows that changes in amplitude of extracted sinusoids are small.

In Figure 4, the differences of the frequency vary. But the amplitudes of the sinusoids are low at high difference range and high at low difference range. This shows that frequency differences of important components in audio signals gather near 0.

Table I shows PSNR and the data rates of speech compressed by the proposed method and G.729. The data rate after the proposed compression was 4.4kbps. The PSNR of the band passed audio signal and the compressed and decoded signal were 23.38db and 24.17db, respectively. The audio quality was not good. The content can be heard, but the audio quality was worse than telephone and the plosive sound can not be heard. This is due to the band-pass filter. Compared to the G.729, the PSNR is lower but it can not be compared fairly due to the difference of the data rates. This will be our future work.

VI. Conclusion

A method to extract long consecutive sinusoids in multiple consecutive frames in the target signal using GHA is proposed. Sinusoids which is correspond to the musical score are extracted from an audio signal of a popular music.

Changes in amplitude and frequency of extracted sinusoids are small. A digital audio compression method using this characteristics is also proposed. In the experiment of audio compression, a speech audio was compressed and PSNR was evaluated.

References

[1] N. Wiener, "Generalized harmonic analysis," *Acta Mathematica*, vol. 55, pp. 117–285, 1930. [Online]. Available: http://ci.nii.ac.jp/naid/10016594318/en/

[2] Y. Hirata and T. Koike, "Speech band compression using a generalized harmonic analysis," *Technical report of IEICE. EA*, vol. 98, no. 277, pp. 17–24, 19980918. [Online]. Available: http://ci.nii.ac.jp/naid/110003284392/en/

[3] M. Nakazawa and Y. Yamasaki, "Sound coding using 1/12n octave analysis," *GITS/GITI research bulletin*, vol. 2002, pp. 81–85, 20030731. [Online]. Available: http://ci.nii.ac.jp/naid/110000973690/en/

[4] E. B. George, "Analysis-by-synthesis/overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones," *J. Audio Eng. Soc.*, vol. 40, no. 6, pp. 497–515, 1992. [Online]. Available: http://ci.nii.ac.jp/naid/80006647890/en/

[5] S. Ushiyama, M. Tohyama, M. Iizuka, and Y. Hirata, "Generalized harmonic analysis of non-stationary waveforms," *Technical report of IEICE. EA*, vol. 93, no. 527, pp. 39–44, 19940322. [Online]. Available: http://ci.nii.ac.jp/naid/110003284275/en/

[6] N. Corp., *NVIDIA CUDA Compute Unified Device Architecture Programming Guide Version 2.0*, 20080607. [Online]. Available: http://www.nvidia.com/object/cuda_home.html

[7] gpgpu.org, "General-purpose computation on gpus (gpgpu)," *http://gpgpu.org/*. [Online]. Available: http://gpgpu.org/