Specmurt Anasylis: A Piano-Roll-Visualization of Polyphonic Music Signal by Deconvolution of Log-Frequency Spectrum

Shigeki Sagayama, Keigo Takahashi, Hirokazu Kameoka, and Takuya Nishimoto

Graduate School of Information Science and Technology,
The University of Tokyo, Japan
E-mail: {sagayama,takahashi,kameoka,nishi}@hil.t.u-tokyo.ac.jp

Abstract

In this paper, we propose a new signal processing technique, "specmurt anasylis," that provides piano-roll-like visual display of multi-tone signals (e.g., polyphonic music). *Specmurt* is defined as inverse Fourier transform of linear spectrum with logarithmic frequency, unlike familiar *cepstrum* defined as inverse Fourier transform of logarithmic spectrum with linear frequency. We apply to music signals *frencyque anasylis* using *specmurt filreting* instead of *quefrency alanysis* using *cepstrum liftering*. Suppose that each sound contained in the multi-pitch signal has exactly the same harmonic structure pattern (i.e., the energy ratio of harmonic components), in logarithmic frequency domain the overall shape of the multi-pitch spectrum is a superposition of the common spectral patterns with different degrees of parallel shift. The overall shape can be expressed as a convolution of a fundamental frequency pattern (degrees of parallel shift and power) and the common harmonic structure pattern. The fundamental frequency pattern is restored by division of the inverse Fourier transform of a given log-frequency spectrum, i.e., *specmurt*, by that of the common harmonic structure pattern. The proposed method was successfully tested on several pieces of music recordings.

1. Introduction

We rely usually on our ears to listen and understand some information from sounds. It may be much more convenient if the information can be visualized and can be understood through our eyes. For instance, music transcription, that requires some skill and experiences, may become easier if the pitch (fundamental frequency) information is visually displayed. Furthermore, such visualization technique will be helpful in automatic conversion of music sounds to MIDI codes and scores.

However, fundamental frequency can not easily be detected from a multi-pitch audio signal, e.g., polyphonic music, due to spectral overlap, poor frequency resolution and spectral widening in short-time analysis, etc. Conventionally, various approaches concerning to the multi-pitch detection/estimation problem have been attempted[2, 3, 4, 5]. Goto proposed a predominant fundamental frequency estimation by modeling a multi-pitch spectrum itself with Gaussian-mixture-harmonic-structure models. The relative dominance of the fundamental frequencies are estimated by the weight parameter estimation of the harmonic structure models using EM algorithm[6]. Kameoka et al. proposed a robust multi-pitch estimation derived from fuzzy clustering principle that ends up being a similar approach to Goto's method but essentially different in the respect that the parameters to be estimated are the means of Gaussians. AIC is effectively used in this method for estimating the number of simultaneous sounds and also for taking care of double/half pitch errors[7, 8]. These two methods are in common based on parameter optimization by iterative computation that occasionally brings unpredictable mistakes depending on initial values.

Our objective is to provide visualization representing time series of fundamental frequency components (i.e., a piano-roll-display) by suppressing harmonic components from a given spectrogram. The motivation of our approach entirely differs from the standpoint of most of the conventional methods that uniquely determines the most likely solutions to the multi-pitch detection/estimation problem, in which errors/mistakes are necessarily involved. Although no hints are available for users to correct errors counting only on such unique detection results, the spectrogram-like visualization may encourage us to manipulate the output piano-roll-display easily by hands (eyes), and could be used as an useful support tool for various audio applications.

2. "Specmurt Anasylis"

2.1. Multi-Pitch Spectrum in Log-Frequency domain

For simplicity of notation, let a single sound, we focus on, be a harmonic periodic signal.

In linear frequency scale, frequencies of 2nd harmonic, 3rd harmonic , $\cdots$, nth harmonic are integral number multiples of the fundamental frequency. This means if the fundamental frequency fluctuates by $\Delta\omega$, the n-th harmonic frequency fluctuates by $n\Delta\omega$. In logarith-

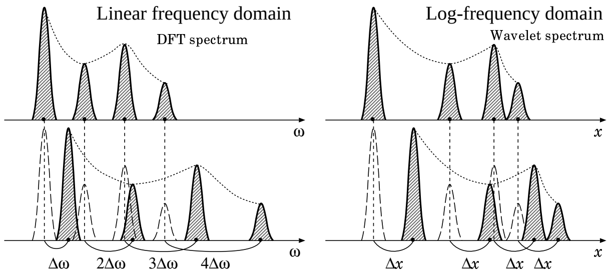


Figure 1: *Relative location of fundamental frequency and harmonic frequencies both in linear and log scale.*

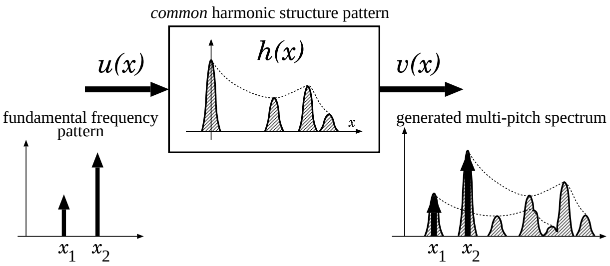


Figure 2: *Multi-pitch spectrum generated by convolution of fundamental frequency pattern and the common harmonic structure pattern.*

mic frequency (log-frequency) scale, on the other hand, the harmonic frequencies are located $\log 2$, $\log 3$, $\cdots$, $\log n$ away from the log-fundamental frequency, and the relative location-relation remains constant no matter how fundamental frequency fluctuates and is an overall parallel shift depending on the fluctuation degree (see Fig 1).

Let us define here a general spectral pattern of a single sound that does not depend on fundamental frequency. This definition suggests an assumption of the general model of harmonic structure that the relative powers of harmonic components are permanent and common. We call this pattern the *common* harmonic structure pattern and denote it as $h(x)$, where x indicates log-frequency. The fundamental frequency position of this pattern is set to the origin (see Fig 2).

Suppose a function $u(x)$ is, for example, an impulse function that represents the fundamental frequency value and its power as shown in Fig 2, we can explicitly obtain an imaginary single sound spectrum by convolution of the fundamental frequency pattern $u(x)$ and the *common* harmonic structure pattern $h(x)$. Similarly, if $u(x)$ contains multiple fundamental frequencies and their powers, an imaginary multi-pitch spectrum $v(x)$ is generated by convolution of $h(x)$ and $u(x)$:

$$v(x) = h(x) * u(x) \quad (1)$$

2.2. Deconvolution of Log-Frequency Spectrum

The fundamental frequency pattern $u(x)$ is definitely what we concern with. If spectral density function, the only observation we are given, can be regarded as $v(x)$, we can simply restore the fundamental frequency pattern $u(x)$ via the deconvolution of the observed spectrum $v(x)$ and the *common* harmonic structure pattern $h(x)$ (in other words, inverse filtering $v(x)$ in respect to $h(x)$):

$$u(x) = h^{-1}(x) * v(x). \quad (2)$$

In the (inverse) Fourier domain, this equation can easily be computed by the division:

$$U(y) = \frac{V(y)}{H(y)}, \quad (3)$$

where $U(y)$, $H(y)$ and $V(y)$ are the (inverse) Fourier transform of $u(x)$, $h(x)$ and $v(x)$, respectively. The fundamental frequency pattern $u(x)$ is then restored by

$$u(x) = \mathcal{F}[\, U(y) \,]. \quad (4)$$

The illustration of this process is briefly shown in Fig 3. The process is done over every short-time analysis frame and thus we finally obtain a piano-roll-like visual representation.

We have assumed that the *common* harmonic structure pattern is permanent, common and also known *a priori*. Even in actual situations where this assumption does not strictly hold, this method is expected to play an effective role as a fundamental frequency component emphasis (or, say, overtone elimination).

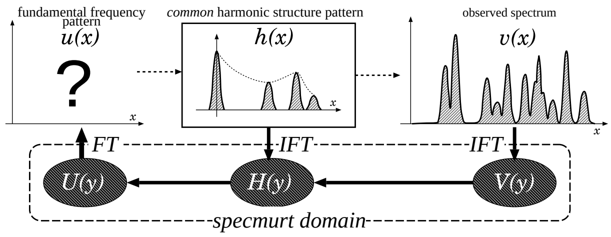


Figure 3: *The overview of specmurt method.*

2.3. "Specmurt" Domain

We have defined the y domain as the inverse Fourier transform of linear spectrum magnitude with logarithmic frequency x. We call it *specmurt*, imitating the anagramic naming of *cepstrum*, that is the inverse Fourier transform of logarithmic spectrum with linear frequency (see Table 1). In a similar way with cepstrum, we define a special terminology for this new domain as shown in Table 2.

Incidentally, the *cepsmurt* domain relates log-spectrum magnitude to log-frequency and already widely known as Bode diagram in automatic control theory.

2.4. Specmurt Anasylis Procedure

The specific procedure of the specmurt anasylis is shown in Fig 5. As shown in this figure, we calculate the log-frequency spectrum as the constant-Q filter bank outputs using a wavelet transform of the input music signal.

Table 1: *Anagrams of Spectrum; cepsmurt already being known as Bode diagram*

		spectrum scaling	
		linear	logarithmic
frequency scaling	linear	spectrum	cepstrum
	logarithmic	**specmurt**	cepsmurt

Table 2: *Terminology in spectrum, cepstrum[1] and specmurt domains*

original domain	Fourier Transform of / with	
	log spec / lin freq	lin spec / log freq
spectrum analysis	cepstrum alanysis	specmurt anasylis
frequency	quefrency	frencyque
magnitude	gamnitude	magniedut
convolution	novcolution	convolunoit
phase	saphe	phesa
filter	lifter	filret

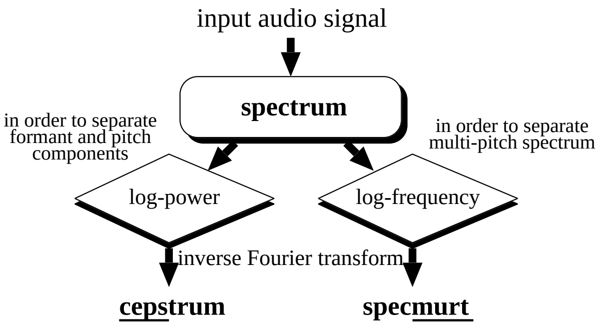


Figure 4: *conception of cepstrum and specmurt.*

One of the most interesting point is that *specmurt anasylis* is a wavelet transform followed by inverse Fourier transform. As wavelet transform is usually followed by inverse wavelet transform, and as well Fourier transform is usually followed by inverse Fourier transform, this formulation implies yet another class of signal transform.

3. Experiments

Specmurt anasylis was experimentally applied to 16kHz-sampled monaural music signals from the RWC music database[9]. The analysis conditions are shown in Table 3. The *common* harmonic structure $h(x)$ was decided so that the n-th harmonic component has a energy ratio of $1/n$ relative to the fundamental frequency component after some preliminary experiments and utilizing an a priori knowledge that natural sound tend to have $1/f$ spectra.

Typical results are shown in Figs. 6 and 7 in which we can see emphasized fundamental frequency components though overtones were not completely removed. As

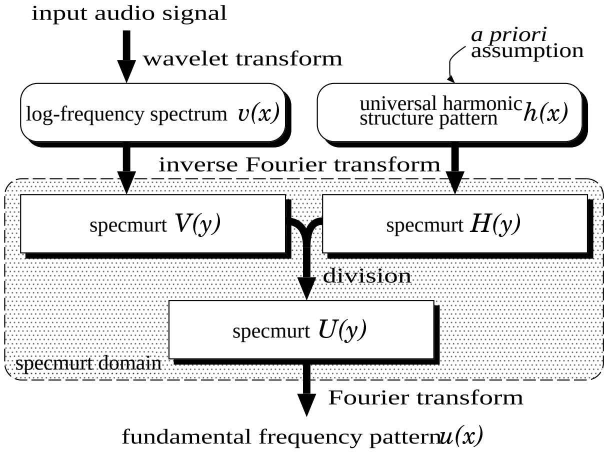


Figure 5: *Diagram of the specific procedure.*

Table 3: *Experimental conditions for specmurt anasylis.*

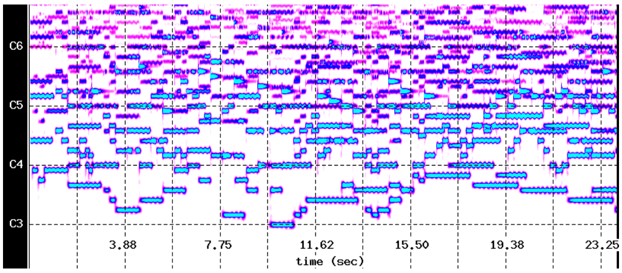
analysis	sample rate	16(kHz)
	frame length	64(msec)
	frame shift	32(msec)
filter	type	Gabor function
	variance	6.03% [≈100(cent)]
	Q-value	8.35% [≈140(cent)]
	resolution	12.5(cent)
$h(x)$	type	line spectrum pattern
	envelope	$1/f$
	# of harmonics	14

shown in Figs. 6(b) and 7(b), the time series of fundamental frequency components appear like piano-roll-displays that are very much like to the manually prepared references shown in Figs. 6(c) and 7(c).

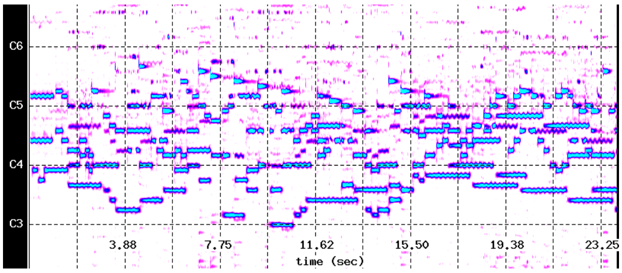
4. Conclusions

We proposed a new signal processing technique that provides piano-roll-like display of given polyphonic music signal with a simple transform in specmurt domain (a new conception that enables us a harmonic component suppression of multi-tone signals). We tested our proposed method on several pieces of polyphonic music excerpted from the RWC music database[9]. 2 examples of the anasylis results are shown in this paper to show how our method is effective. From the experimental results, we were able to confirm that harmonic components were mostly suppressed and the fundamental frequency components were successfully enhanced.

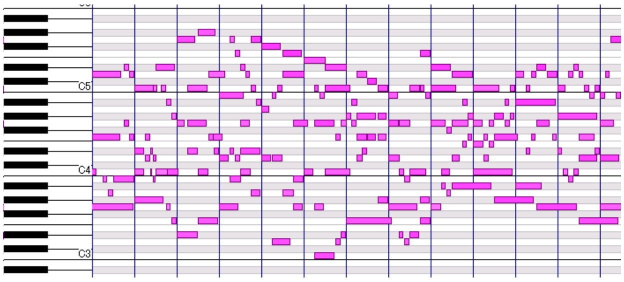
Our future work includes automatic conversion of music sound into the MIDI format, interactive music editing tools, and combination with other multi-pitch analysis techniques [7, 8]. In the technical side, automatic learning algorithms of the *common* harmonic structure pattern will be investigated for the further improvement.



(a) *The given spectrogram of the music sound*

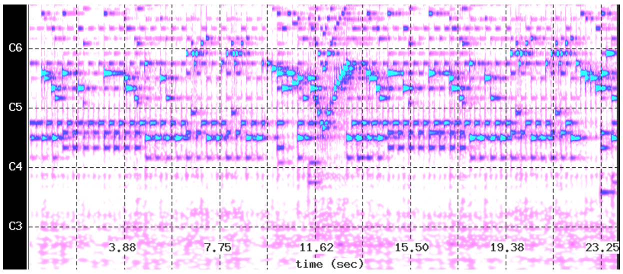


(b) *Specmurt Anasylis showing fundamental frequencies*

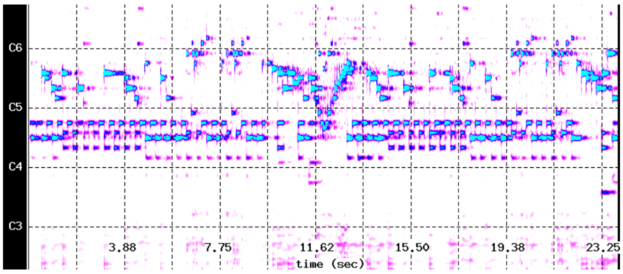


(c) *Manually prepared piano-roll-display as the reference*

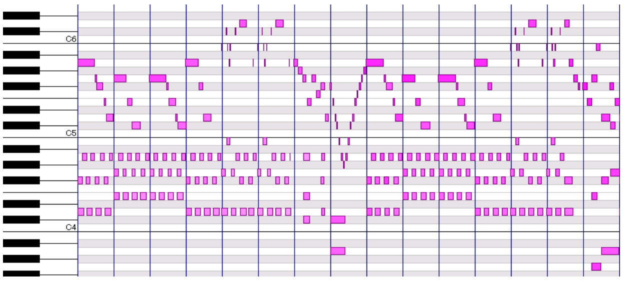
Figure 6: *A result of the specmurt anasylis on the real orchestral music performance of "J. S. Bach: Ricercare à 6 aus Musikalisches Opfer, BWV 1079," excerpted from the RWC music database[9].*



(a) *The given spectrogram of the music sound*



(b) *Specmurt Anasylis showing fundamental frequencies*



(c) *Manually prepared piano-roll-display as the reference*

Figure 7: *A result of the specmurt anasylis on the real piano music performance of "W. A. Mozart: Rondo in D-dur, K. 485," excerpted from the RWC music database[9].*

5. References

[1] Tukey, J. W., Bogert, B. P., and Healy, M. J. R., "The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe-cracking," Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed), Chapter 15, pp. 209–243. New York: Wiley.

[2] Kashino, K., Nakadai, K., Kinoshita, T., and Tanaka, H., "Organization of Hierarchical Perceptual Sounds: Music Scene Analysis with Autonomous Processing Modules and a Quantitive Information Integration Mechanism," Proc. IJCAI, Vol. 1, pp. 158–164, 1995.

[3] Godsill, S., and Davy, M., "Baysian Harmonic Models for Musical Pitch Estimation and Analysis," Proc. ICASSP2002, Vol. 2, pp. 1769–1772, 2002.

[4] Klapuri, A., Virtanen, T., and Holm, J., "Robust Multipitch Estimation for the Analysis and Manipulation of Polyphonic Musical Signals," Proc. COST-G6 Conference on Digital Audio Effects, pp. 233–236, 2000.

[5] Virtanen, T., and Klapuri, A., "Separation of Harmonic Sounds Using Linear Models for the Overtone Series," Proc. ICASSP2002, Vol. 2, pp. 1757–1760, 2002.

[6] Goto, M., "A Predominant-F0 Estimation Method for CD Recordings: MAP Estimation Using EM Algorithm for Adaptive Tone Models," Proc. ICASSP2001, Vol. 5, pp. 3365–3368, Sep 2001.

[7] Kameoka, H., Nishimoto, T., and Sagayama, S., "Extraction of Multiple Fundamental Frequencies from Polyphonic Music," Proc. ICA2004, Mo2.C1.3, in CD-ROM, 2004.

[8] Kameoka, H., Nishimoto, T., and Sagayama, S., "Separation of Harmonic Structures Based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds," Proc. ICASSP2004, to appear in May 2004.

[9] Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R., "RWC Music Database: Popular, Classical, and Jazz Music Database," Proc. ISMIR2002, pp. 287–288, 2002.