Specmurt Anasylis: A Piano-Roll-Visualization of Polyphonic Music Signals 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 this technique 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

Detecting and estimating multiple fundamental frequencies is essential for automatic/semi-automatic music transcription, conversion to MIDI signals, music information retrieval, etc. However, fundamental frequency can not easily be detected from a multi-pitch audio signals such as multitone or polyphonic music, mainly due to spectral overlap, poor frequency resolution and widened spectrum in short-time analysis, etc. Conventionally, various approaches concerning the multi-pitch detection/estimation problem have been attempted[2, 3, 4, 5]. Goto[6] 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 the EM algorithm. Kameoka et al.[7, 8] proposed a robust multi-pitch estimation derived from fuzzy clustering principle similar to Goto's approach but different in 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. These two methods are commonly based on parameter optimization by iterative computation that occasionally brings unpredictable mistakes depending on initial values.

Our objective is to provide a visualization technique representing fundamental frequency components by suppressing harmonic components in the given spectrum and produce a "piano-roll" display similar to that of MIDI signal display. The motivation of our approach is different from those of most conventional methods that give only the most likely solutions to the multi-pitch detection/estimation problem, in which errors/mistakes are necessarily involved partly due to local optimum problems. Instead, the visualization approach gives a global image with "soft decision" of fundamental frequencies of the signal. The result can be used in automatic or semi-automatic transcription of music, conversion of music signal into MIDI format, and efficient initial value estimation for more precise multipitch analysis[7, 8].

2. "Specmurt Anasylis"

2.1. Multi-Pitch Spectrum in Log-Frequency domain

First, we discuss a single-tone signal with a single fundamental frequency and a harmonic structure. In the 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 $n\Delta\omega$. On the other hand, in the logarithmic frequency (log-frequency) scale, 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 assume that all single-tone signals have a common harmonic structure which does not depend on the fundamental frequency. We call it the *common* harmonic structure and denote it as $h(x)$, where x represents the

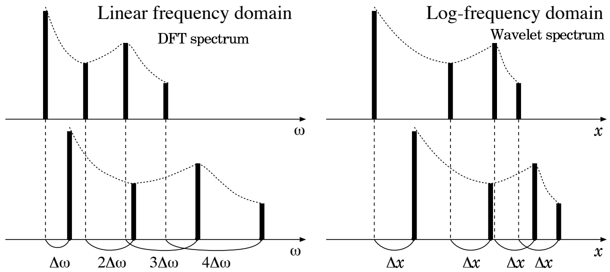


Figure 1: *Relative location of fundamental frequency and harmonic frequencies in linear and logarithmic scales.*

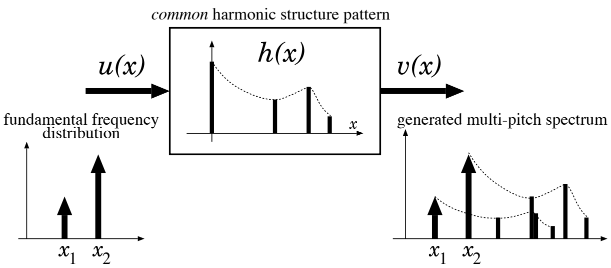


Figure 2: *Multi-pitch spectrum generated by convolution of fundamental frequency pattern and the common harmonic structure pattern.*

logarithmic frequency. The fundamental frequency position of this pattern is set to the origin (see Fig 2). Obviously, this assumption is not true for real music sounds, but is practically approximate in many cases as shown later.

Next, we define a function $u(x)$ to represent the distribution of fundamental frequencies in a multipitch signal. If $u(x)$ is simply an impulse function, for instance, it represents the logarithmic fundamental frequency and the power of the single tone with a harmonic structure $h(x)$.

If we assume that the power spectrum is additive[1], the power spectrum of a multipitch signal is represented as a convolution of the fundamental frequency distribution $u(x)$ and the *common* harmonic structure $h(x)$:

$$v(x) = h(x) * u(x) \tag{1}$$

as shown in Fig 2. In other words, the power spectrum $v(x)$ of a multipitch signal can be regarded as the output of a filter $h(x)$ representing the common harmonic structure given the input $u(x)$ representing the fundamental frequency distribution. This relation can be extended to non-harmonic and/or continuous spectrum $h(x)$ and continuous distribution $u(x)$ if the single-tone spectrum is log-frequency shift-invariant and power spectrum is additive.

[1] This is always true only in the expectation sense. The power spectrum of sum of two signals is not exactly equal to the sum of two spectra of the signals and depends on phase difference between components of the same frequency in the two signals. But, this assumption is widely accepted, e.g., in spectrum subtraction for noise reduction from noisy speech signals.

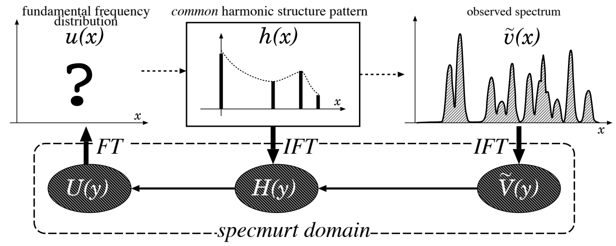


Figure 3: *The overview of specmurt method.*

2.2. Deconvolution of Log-Frequency Spectrum

Regarding $v(x)$ as the observed power spectral density function of a multipitch signal, the fundamental frequency distribution $u(x)$ can be restored via the deconvolution of the observed spectrum $v(x)$ with the *common* harmonic structure pattern $h(x)$ (in other words, inverse filtering $v(x)$ in respect to $h(x)$):

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

In the (inverse) Fourier domain, this relation is written as a division:

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

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

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

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

2.3. "Specmurt" Domain

We defined $V(y)$ as the inverse Fourier transform of linear power spectrum $v(x)$ with logarithmic frequency x. We call it *specmurt*, imitating the anagramic naming of *cepstrum*[1], that is the inverse Fourier transform of logarithmic spectrum with linear frequency (see Table 1)[2] and called "*quefrency alanysis*". In the same way, as cepstrum, a special terminology for this new domain can be defined as shown in Table 2.

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.

One 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 Fourier transform by inverse

[2] *Cepsmurt* at the bottom right in Table 1 is Fourier transform of log-spectrum as a function of log-frequency, already well-known as "Bode diagram" in automatic control theory. *Cepsmurt* is close to mel-cepstrum widely used in speech recognition. Note that spectrum is left linear in the *specmurt* case.

Table 1: *Anagrams of Spectrum*

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

Table 2: *Terminology in spectrum, cepstrum[1] and specmurt domains; lefthandside anagrams were defined in [1]*

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

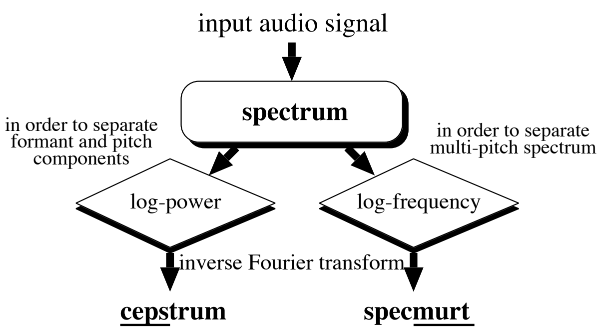


Figure 4: *conception of cepstrum and specmurt.*

Fourier transform, this new pairing implies a new class of signal transform.

We have assumed that the harmonic structure $h(x)$ is common, constant over time, and also known *a priori*. Even if this assumption does not strictly hold in actual situations, this method is expected to effectively emphasize the fundamental frequency components and suppress overtones. In practice, $h(x)$ is given heuristically, experimentally or recursively estimated to minimize the residual overtone energy[10].

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 determined 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

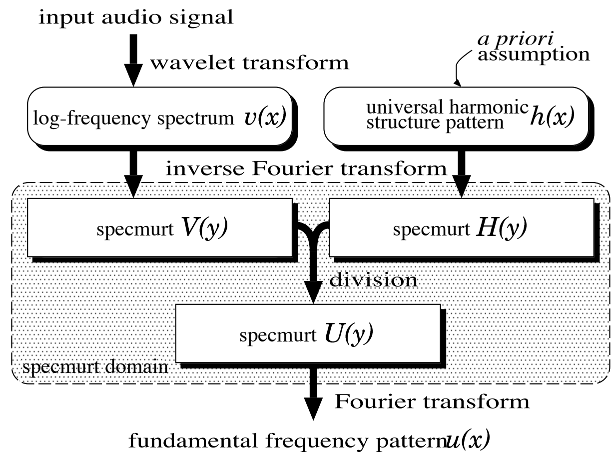


Figure 5: *Diagram of the specific procedure.*

Table 3: *Experimental conditions for specmurt anasylis.*

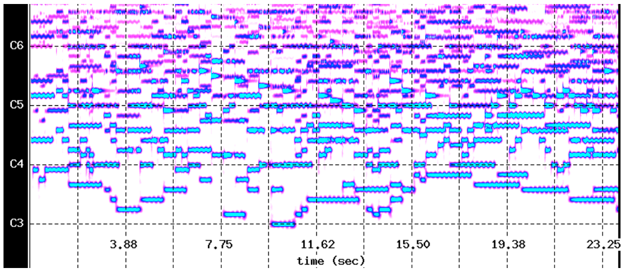
analysis	sample rate	16(kHz)
	frame length	64(msec)
	frame shift	32(msec)
filter	type	Gabor function
	variance	6.03% [$\approx$100(cent)]
	Q-value	8.35% [$\approx$140(cent)]
	resolution	12.5(cent)
$h(x)$	type	line spectrum pattern
	envelope	$1/f$
	# of harmonics	14

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).

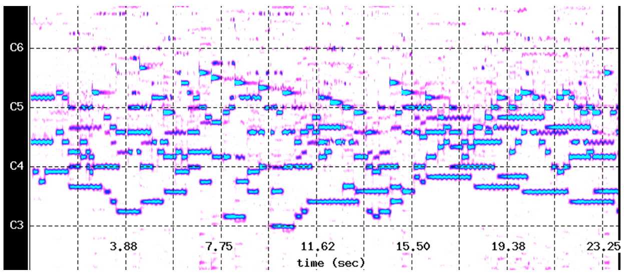
4. Conclusions

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.

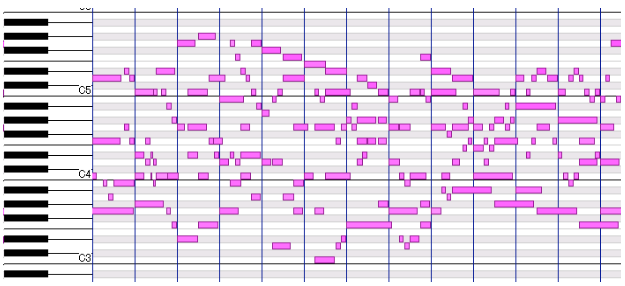
Our future work includes automatic conversion of music sound into the MIDI format, interactive music editing tools, and combination with other multi-pitch analysis techniques [7, 8]. In the technical side, automatic learning algorithms of the *common* harmonic structure pattern will be investigated for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



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

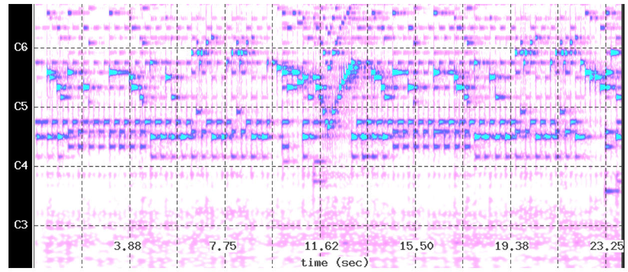


(b) *Specmurt Anasylis showing fundamental frequencies*

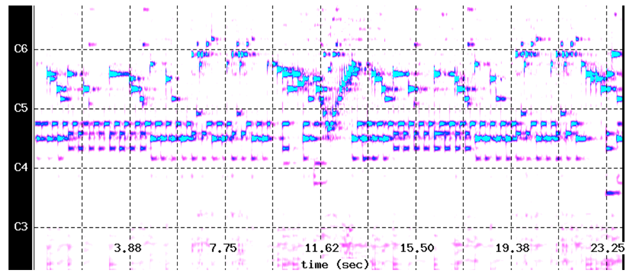


(c) *Manually prepared piano-roll-display as the reference*

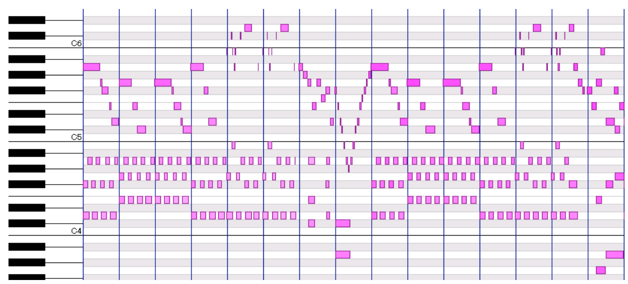
Figure 6: *A result of the specmurt anasylis on the real orchestral music performance of "J. S. Bach: Ricercare à 6 aus Musikalisches Opfer, BWV 1079," excerpted from the RWC music database[9].*



(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.

[10] Kameoka, H., Saito, S., Nishimoto, T., and Sagayama, S., "Recursive Estimation of Quasi-Optimal Common Harmonic Structure Pattern for Specmurt Anasylis: Piano-Roll-Display Visiualization and MIDI Conversion of Polyphonic Music Signal," IPSJ Technical Report, SIG-MUS, to appear in August, 2004. (in Japanese)