‘Specmurt Anasylis’ of Multi-Pitch Signals

Shigeki Sagayama, Hirokazu Kameoka, Shoichiro Saito and Takuya Nishimoto
The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8656 Japan
E-mail: {sagayama,kameoka,saito,nishi}@hil.t.u-tokyo.ac.jp

***Abstract*— In this paper, we discuss a new concept of *specmurt* to analyze multi-pitch signals. It is applied to polyphonic music signal to extract fundamental frequencies, to produce a piano-roll-like display, and to convert the saound into MIDI data.**

In contrast with *cepstrum* which is the inverse Fourier transform of log-scaled power spectrum with linear frequency, *specmurt* is defined as the inverse Fourier transform of linear power spectrum with log-scaled frequency. If all tones in a polyphonic sound have a common harmonic pattern, it can be regarded as a sum of frequency-stretched common harmonic structure. In the log-frequency domain, it is formulated as the convolution of distribution density of fundamental frequencies of multiple tones and the common harmonic structure. The fundamental frequency distribution can be found by deconvolution, i.e., by division in the *specmurt* domain.

This ‘specmurt anasylis’ is demonstrated in generation of a piano-roll-like display from a polyphonic music signal and in automatic sound-to-MIDI conversion.

I. Introduction

In 1963, Bogert, Healy and Tukey introduced a concept of ‘cepstrum’ in a paper entitled “*The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe-cracking*” [1]. Later on, this basic idea was widely applied in signal processing not limited within analysis of echos. Since Noll [2] used cepstrum in pitch detection in 1964, it became a standard technique for determination of fundamental frequency. In speech recognition, Kohda, Nagashima and Sagayama worked out cepstrum-based speech recognition at NTT Labs in 1978, though this fact was not officially recorded, and used in the ANSER system later, the first deployment of speech recognition in public service. Together with delta-cepstrum proposed by Sagayama and Itakura in 1979 [3], cepstrum has been the speech feature most often used in speech recognition in the form of ‘mel-frequency cepstrum coefficients’ (MFCC) proposed by Davis and Mermelstein in 1980. Cepstrum was also utilized in speech synthesis digital filter by Imai and Kitamura in 1978 [5].

In these applications, cepstrum is advantageous in converting the speech spectrum into a sum of pitch and envelope components in the cepstrum domain. It is assumed, however, that cepstrum treats a single signal. Multi-pitch signal can not be handled by cepstrum due to its non-linearity of logarithm.

Multi-pitch analysis has been one of major problems in music sound signal processing. However, fundamental frequency can not easily be detected from a multi-pitch audio signal, e.g., polyphonic music, due to spectral overlap of overtones, poor frequency resolution and spectral widening in short-time analysis, etc. Conventionally, various approaches concerning the multi-pitch detection/estimation problem have been attempted [8], [9], [10], [11], [12], [13]. Reliable determination of the number of sound sources is discussed only recently [14].

As for spectral analysis, wavelet transform using the Gabor function is one of the popular approach to derive short-time power spectrum of music signals along logarithmically-scaled frequency axis that appropriately suits the music pitch scaling. Spectrogram, i.e., a 2-dimensional time-frequency display of the sequence of short-time spectra, however, is apparently messed up because of the existence of many overtones (i.e., the harmonic components of multiple fundamental frequencies), that often prevents us from discovering music notes.

Our objective is to emphasize the fundamental frequency components by suppressing the harmonic components so that the spectrogram will become more similar to the piano-roll display from which we can see multiple fundamental frequencies in the display. The motivation of our approach entirely differs from the standard multi-pitch analysis methods that uniquely determines the most likely solutions to the multi-pitch detection/estimation problem, in which errors are necessarily involved. This kind of errors are often unpredictable(e.g., recursive solutions depends highly on initial values), that could be harmful when simply using the detection results for a music retrieval purpose. The ‘Specmurt Anasylis’, on the other hand, provides visually similar display to the original piano-roll image, that may hopefully be one of the useful features for the retrieval purpose(imagine a simple image template matching for instance).

As for single pitch detection and extraction, the well-known cepstrum is the inverse Fourier transform of log-scaled spectrum along linear frequency axis. In contrast, we use *specmurt* that is the inverse Fourier transform of linear-scaled spectrum along log-frequency axis. The proposed method was successfully tested on several pieces of music recordings.

II. ‘Cepstrum’ versus ‘Specmurt’

A. ‘Cepstrum Alanysis’

According to Wiener-Khinchin Theorem, the inverse Fourier transform of linear power spectrum is autocorrelation as a function of time delay as follows:

$$v(\tau) = \int_{-\infty}^{\infty} e^{j\tau\omega} f(\omega) d\omega, \quad -\infty < \tau < \infty \tag{1}$$

where $f(\omega)$ denotes the power spectrum of the signal. If power spectrum is scaled logarithmically, the resulted inverse Fourier transform is not autocorrelation any more and is named ‘cepstrum’ [1], humorously reversing the first four letters in

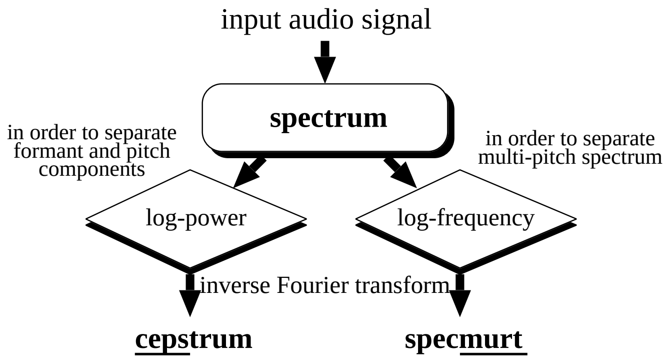


Fig. 1. Contrast between cepstrum and *specmurt* : *specmurt* is defined as inverse Fourier transform of linear spectrum with log-frequency, whereas cepstrum is inverse Fourier transform of log spectrum with linear frequency.

the spelling of 'spectrum.' It is formulated as follows:

$$c(q) = \int_{-\infty}^{\infty} e^{jq\omega} \log f(\omega) d\omega, \quad -\infty < q < \infty \qquad (2)$$

where q is called 'quefrency' and usually chosen to be an integer for band-limited spectra.

Analyzing signals in the cepstrum domain is referred to 'quefrency alanysis' instead of 'frequency analysis.' Similarly, derivatives of inverse Fourier transform of log-scaled power spectrum are named 'gamnitude', 'novcolution', 'saphe' and 'lifter' by partially reversing the spellings of existing words in spectrum analysis: magnitude, convolution, phase and filter.

If a signal is produced by a periodic excitation signal of a single pitch frequency convolved with an impulse response of linear filter forming the spectrum envelope, they are often expected to be separate in the quefrency domain. It is useful in single pitch analysis.

B. *'Specmurt Anasylis'*

Instead of inverse Fourier transform of log-scaled power spectrum with linear frequency in the cepstrum case, we can also consider inverse Fourier transform of linear power spectrum with log-scaled frequency as follows:

$$s(y) = \int_{-\infty}^{\infty} e^{jy\log\omega} f(\omega) d\log\omega, \quad -\infty < y < \infty \qquad (3)$$

or, denoting $x = \log\omega$ and $g(x) = f(\omega)$:

$$s(y) = \int_{-\infty}^{\infty} e^{jxy} g(x) dx, \quad -\infty < y < \infty \qquad (4)$$

which we call *specmurt* [6] reversing the last four letters in the spelling of 'spectrum' respecting the terminology of cepstrum. Signal analysis in the *specmurt* domain is referred to 'specmurt anasylis' instead of 'spectrum analysis' and 'cepstrum alanysis.' Specmurt is a function of 'frencyque' y instead of 'frequency' and 'quenfrency.' Manupilation in the *specmurt* domain is referred to 'filret' instead of 'filter' in the spectrum domain and 'lifter' in the cepstrum domain.

In the next section, 'specmurt anasylis' is shown to be effective in multi-pitch signal analysis in contrast with cepsrum for the single-pitch case.

TABLE I
TERMINOLOGY IN SPECTRUM, CEPSTRUM[1] AND *specmurt* DOMAINS

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

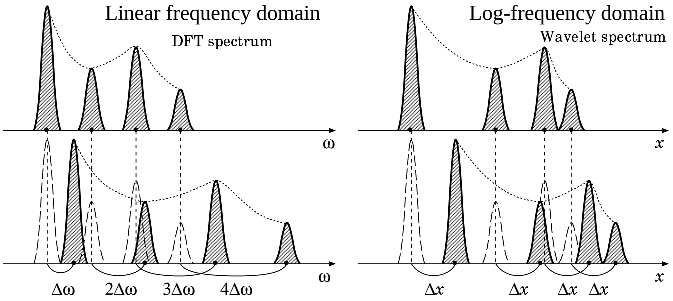


Fig. 2. Relative location of fundamental frequency and harmonic frequencies both in linear and log scale.

It should be noted that spectrum logarithmically scaled both in frequency and in magnitude is identical to Bode diagram often used in the automatic control theory. Its Fourier transform has no specific name, while it is essentially similar to mel-scaled frequency cepstrum coefficients (MFCC) and is very often used in the feature analysis in speech recognition.

III. DECONVOLUTION OF LOG-FREQUENCY SPECTRUM

A. *Modeling Single-Pitch Spectrum in Log-Frequency Domain*

For simplicity, we assume that a single sound component is 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 changes by $\Delta\omega$, the n-th harmonic frequency changes by $n\Delta\omega$. In the logarithmic frequency (log-frequency) scale, on the other hand, the harmonic frequencies are located $\log 2$, $\log 3$, $\cdots$, $\log n$ away from the fundamental log-frequency, and the relative location-relation remains constant no matter how fundamental frequency changes and is an overall parallel shift depending on the change (see Fig 2).

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 common. We call this pattern the *common harmonic structure* 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 3).

Suppose a function $u(x)$ is, for example, an impulse (Dirac's delta-function) that represents the fundamental frequency position on the x-axis and the energy of the fundamental frequency

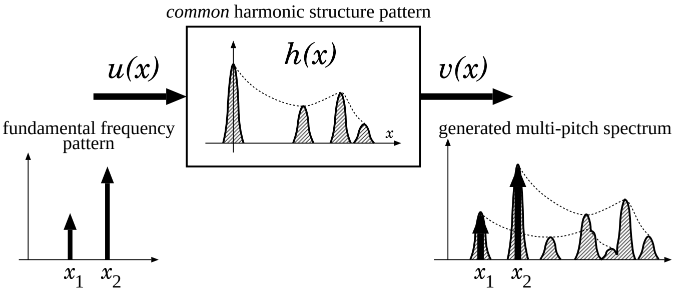

common harmonic structure pattern
u(x)
h(x)
v(x)
x
fundamental frequency pattern
generated multi-pitch spectrum
x_1 x_2
x_1 x_2

Fig. 3. Multi-pitch spectrum generated by convolution of fundamental frequency pattern and the common harmonic structure pattern.

component, we can explicitly obtain a single sound spectrum by convolving the fundamental frequency location $u(x)$ and the common harmonic structure $h(x)$.

B. Modeling Multi-Pitch Spectrum in Log-Frequency Domain

If $u(x)$ contains multiple fundamental frequencies and their powers as shown in Fig. 3, the multi-pitch spectrum $v(x)$ is generated by convolution of $h(x)$ and $u(x)$:

$$v(x) = h(x) * u(x) \tag{5}$$

if power spectrum can be assumed additive. (This assumption holds only in the expectation sense; the power of the sum of multiple sinusoids of the same frequency may deviate from the sum of their powers due to their relative phase relationship.)

Eq. 5 also holds if $u(x)$ is a continuous function representing the distribution of fundamental frequencies.

C. Deconvolution of Log-Frequency Spectrum

The main problem here is to estimate the fundamental frequency pattern $u(x)$ from the observed spectrogram $v(x)$. If $h(x)$ is known, we can restore $u(x)$ by applying the inverse filter $h^{-1}(x)$ to $v(x)$. It is deconvolution of the observed spectrum $v(x)$ with the *common* harmonic structure pattern $h(x)$:

$$u(x) = h^{-1}(x) * v(x). \tag{6}$$

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

$$U(y) = \frac{V(y)}{H(y)}, \tag{7}$$

where $U(y)$, $H(y)$ and $V(y)$ are the (inverse) Fourier transform of $u(x)$, $h(x)$ and $v(x)$, respectively. This operation is referred to *specmurt filretting* according to the terminology in Table I. The fundamental frequency pattern $u(x)$ is then restored by

$$u(x) = \mathcal{F}^{-1}[\ U(y)\ ]. \tag{8}$$

The illustration of this process is briefly shown in Fig 4. The process is done over every short-time analysis frame and thus we finally have a time series of fundamental frequency components, i.e., a piano-roll-like visual representation with a small amount of computation.

TABLE II
ANALYSIS CONDITIONS FOR LOG-FREQUENCY SPECTROGRAM.

analysis	sampling 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)

The y domain has been defined as the inverse Fourier transform of linear spectrum magnitude with logarithmic frequency x.

We have discussed so far on the premise of using the *common* harmonic structure pattern that is common over all constituent tones and also known *a priori*. Even in the actual situations where this assumption may not strictly hold, this approach is still expected to play an effective role as a fundamental frequency component emphasis (or, say, overtone suppression).

IV. OPTIMIZATION OF COMMON HARMONIC STRUCTURE

A. Wavelet Transform of Input Signal

We use wavelet analysis using Gabor kernel function, as it provides short-time power spectrum with a constant resolution along the log-frequency axis. It can be understood as constant-Q filter bank analysis along the log-scaled frequency axis and is well suited for the musical pitch scale.

Fig. 6(a) shows an example of wavelet analysis of music sound performed by an orchestra. In this 5-voice portion of J. S. Bach's Ricercare a 6 voci, a flute, 1st and 2nd violins, a viola and violincello are assigned to the 5 voices. Table II lists the analysis conditions. We see that overtones (harmonics components) of individual music tones overlap on each other. It is quite obvious that finding 5 separate melody lines, i.e., trajectories of fundamental frequencies, from this spectrogram is not at all trivial.

Spectrogram (via wavelet transform) of music signal is usually messed up with many overtones(harmonics components) of the individual music notes, that often overlap on each other. If we are hopefully able to remove the overtone components as much as possible from the observed spectrogram, a piano-roll-like visual display of the music signal will be derived, that may be helpful not only in various music applications such as signal-into-MIDI converter or automatic music transcription, but also in music information retrieval.

B. Computational Procedure of 'Specmurt Anasylis'

The procedure of the 'specmurt anasylis' is illustrated in Fig 4. As shown in this figure, the log-frequency spectrum is first computed as the constant-Q filter bank outputs using a wavelet transform of the input music signal. The whole procedure consists of 4 steps as shown below.

Step 1. Apply wavelet transform with Gabor function to the input signal and take the squared absolute values (power-spectrogram magnitudes) $v(x)$ for each frame.

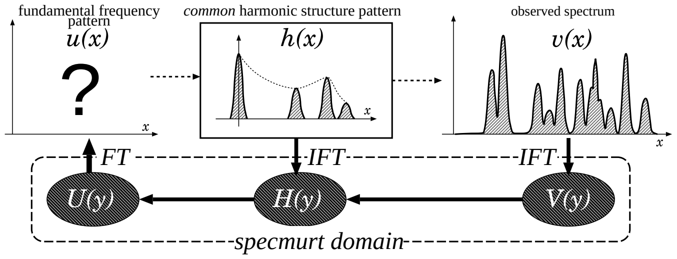


Fig. 4. The outline of the “specmurt anasylis” to find fundamental frequencies.

Step 2. Apply inverse Fourier transform to $v(x)$ to obtain $V(y)$.

Step 3. Divide $V(y)$ by $H(y)$, the inverse Fourier transform of the assumed common harmonic pattern $h(x)$.

Step 4. Fourier transform the division $V(y)/H(y)$ to estimate the multi-pitch distribution $u(x)$ along the log-frequency x.

One interesting aspect of *specmurt anasylis* is that wavelet transform is followed by inverse Fourier transform whereas wavelet transform is usually followed by inverse wavelet transform, or Fourier transform is as well usually followed by inverse Fourier transform.

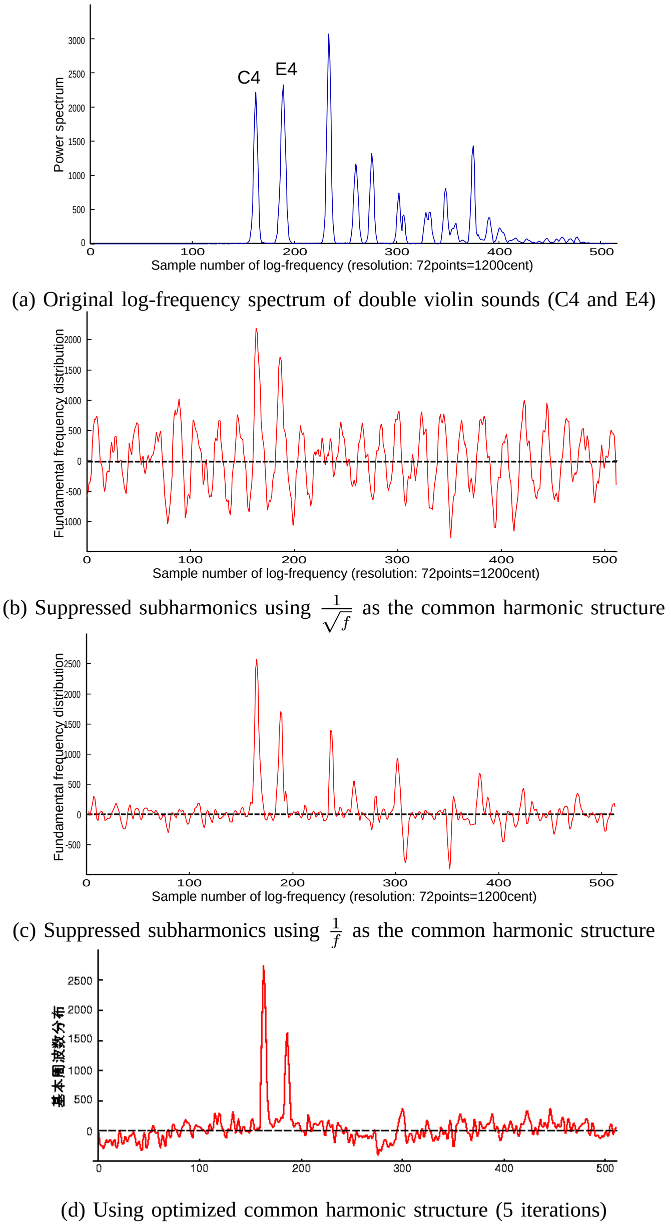
C. Optimizing the Common Harmonic Structure

In the above procedure of ‘specmurt anasylis,’ we assumed that all constituent sounds have a common harmonic structure. It is, however, generally not true in real polyphonic music sounds as the harmonic structures are generally different from each other and they often change over time. The best we can do is to give a best compromise of $h(x)$ to minimize the amplitudes of subharmonics (overtones) after deconvolution (by ‘filretting’) in the *specmurt* domain. This situation is somewhat similar to ceptrum-based pitch extraction where ‘lifter’ should often be empirically adjusted to separate pitch and formant components in the ‘quefrency’ domain.

Figure 5 (a) shows an example of the linear-scaled spectrum of mixture of two violin sounds (C4 and E4) along log-scaled frequency axis x where multiple peaks represent two fundamental frequencies as well as overtones. Using $1/\sqrt{f}$ as the frequency characteristics of $h(x)$ where f denotes frequency, the overtones are suppressed in Figure 5 (b) while unnecessary spectrum components appear as the result of deconvolution. On the other hand, using $1/f$, suppression of overtones is insufficient in Figure 5 (c).

To automatically adjust $h(x)$, we consider an iterative procedure to find the optimal $h(x)$ that gives maximum suppression of subharmonic components. Applying a non-linear mapping to the estimated fundamental frequency distribution, $u(x)$, we can suppress relatively smaller values of $u(x)$ while keeping relatively large values same, to obtain $\bar{u}(x)$ as the mapped result. Then, we can derive $\bar{h}(x)$, an improved $h(x)$ to match $\bar{u}(x)$ in the least squares sense, by minimizing an objective function:

$$D = \int_{-\infty}^{\infty} \left\{ v(x) - \bar{h}(x) * \bar{u}(x) \right\}^2 dx \qquad (9)$$



(a) Original log-frequency spectrum of double violin sounds (C4 and E4)

(b) Suppressed subharmonics using $\frac{1}{\sqrt{f}}$ as the common harmonic structure

(c) Suppressed subharmonics using $\frac{1}{f}$ as the common harmonic structure

(d) Using optimized common harmonic structure (5 iterations)

Fig. 5. Multi-pitch extraction with optimized common harmonic structure.

The procedure of ‘specmurt anasylis’ is repeated to obtain an yet improved $u(x)$ using the improved common harmonic structure $\bar{h}(x)$. The whole procedure is relatively simple as $h(x)$ has non-zero values at its fundamental and harmonic frequencies.

A practical procedure for this purpose is as follows:

Step 1. Obtain $\bar{u}(x)$ by applying a non-linear mapping utilizing a sigmoid function:

$$\bar{u}(x) = \frac{1}{1+\exp\left\{-\alpha\big(u(x)-\beta\big)\right\}} u(x) \qquad (10)$$

where sigmoid parameters α and β are chosen based on the distribution of values of $u(x)$.

Step 2. Find $\bar{h}(x)$ at N discrete points $\{x_1, x_2, \cdots, x_N\}$ (N is the number of subharmonics to consider and

TABLE III
SOUND-TO-MIDI CONVERSION ACCURACY (%)

Title	Instrument	Genre	Composer / Player	Correct Rate(%)
“Jive”	Piano	Jazz	M. Nakamura	77.8
“Lounge Away”	Piano	Jazz	T. Nagai	78.4
“Jive”	Guitar	Jazz	H. Chubachi	77.6
“For Two”	Guitar	Jazz	H. Chubachi	76.9
“Crescent Serenade”	Guitar	Jazz	S. Yamamoto	74.5
“Abyss”	Guitar	Jazz	H. Chubachi	72.0
Nocturne No. 2, E♭ major, op. 9-2	Piano	Classical	F. Chopin	80.4

$x_k = x_1 + \log k$) by calculating $\{h_1, h_2, \cdots, h_N\}$ through the following equations:

$$a_{j,k} = \int_{-\infty}^{\infty} u(x - x_j) u(x - x_k) dx \tag{11}$$

$$b_j = \int_{-\infty}^{\infty} \{v(x) - u(x)\} u(x - x_j) dx \tag{12}$$

$$\begin{pmatrix} a_{1,1} & \cdots & a_{n,1} & \cdots & a_{1,N} \\ \vdots & & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} & \cdots & a_{n,N} \\ \vdots & & \vdots & & \vdots \\ a_{1,1} & \cdots & a_{i,1} & \cdots & a_{1,N} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \\ \vdots \\ h_N \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \\ \vdots \\ b_N \end{pmatrix} \tag{13}$$

Step 3. Replace $h(x)$ with $\bar{h}(x)$, repeat the ‘specmurt ansylis’ procedure and go to Step 1.

An example of iterative optimization of common harmonic structure is shown in Fig. 5. After 5 iteration of the above procedure starting from initial $u(x)$ either in Fig. 5(b) or (c), the same converged result was obtained as shown in (d).

V. EXPERIMENTS

A. Visualization of Fundamental Frequencies

Specmurt anasylis was experimentally tested on 16kHz-sampled monaural polyphonic music signals from the RWC music database[15].

An example of the ‘specmurt anasylis’ results are shown in Fig. 6, in which we can see the overlapping overtones in (a) is significantly suppressed by ‘specmurt anasylis’ in (b) and is very much like the manually prepared piano-roll references in (c). In this analysis, the envelope of the common harmonic structure $h(x)$ was assumed to be $1/f$ (the n-th harmonic component has a energy ratio of $1/n$ relative to the fundamental frequency component) following an a priori knowledge that natural sounds tend to have ‘$1/f$’ spectral characteristics.

B. Sound-to-MIDI Conversion

Once the fundamental frequencies are found for each frame, they can be converted to MIDI(Musical Instrument Digital Interface)-format data through quantization of fundamental frequencies into music tone names.

Table III shows the conversion accuracy for several music pieces excerpted from a common music database [15]. The conversion accuracy was calculated by counting differences between the MIDI data and the handcrafted MIDI as the reference associated with the music signal data.

VI. CONCLUSION

We discussed a novel non-linear signal processing technique called ‘specmurt anasylis’ which is parallel to ‘cepstrum alanysis’. In this new domain, multiple fundamental frequencies of polyphonic music signal are detected by ‘filretting’ in the *specmurt* domain and displayed in a piano-roll-like display. An iterative optimization of common harmonic structure was also devised and used in sound-to-MIDI conversion of polyphonic music signals.

Our future work includes generalization of ‘specmurt anasylis’ for complex spectra to define yet another homomorphic signal processing similar to complex-cepstrum-based signal processing, providing initial values for precise multi-pitch analysis based on harmonically-constrained Gaussian mixture models[13], [14], application to automatic transcription of music (sound-to-score conversion) by combining with the rhythm transcription technique[16], music performance analysis tools, and interactive music editing/manipulation tools.

REFERENCES

[1] B. P. Bogert, M.J.R. Healry, J.W. Tukey: “ The quefrency alanysis of time series for echos: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe-cracking, ” Proc. Sympo. on Time Series Analysis, Chapter 15, New York: Wiley, pp. 209-243, 1963.

[2] A. M. Noll, ”Short-time spectrum and ‘cepstrum’ techniques for vocal-pitch detection,” J. Acoust. Soc. Amer., vol. 36, no. 2, pp. 296–302, Feb. 1964.

[3] S. Sagayama and F. Itakura, “On Individuality in a Dynamic Measure of Speech,” Proc. ASJ Conf., pp. 589–590, July 1979. (in Japanese)

[4] S. E. Davis and P. Mermelstein, “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences,” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, no. 4, pp. 357–366, Aug. 1980.

[5] S. Imai and T. Kitamura, “Speech Analysis Synthesis System Using Log Magnitude Approximation Filter,” Trans. IEICE Japan, Vol. J61-A, no. 6, pp. 527–534, 1978. (in Japanese)

[6] K. Takahashi, T. Nishimoto and S. Sagayama, “Multi-Pitch Analysis Using Deconvolution of Log-frequency Spectrum,” IPSJ Technical Report, 2003-MUS-53, pp. 61–66, 2003. (in Japanese)

[7] S. Sagayama, H. Kameoka, T. Nishimoto, “Specmurt Anasylis: A Piano-Roll-Visualization of Polyphonic Music Signal by Deconvolution of Log-Frequency Spectrum,” Proc. 2004 ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA2004), Oct. 2004.

[8] K. Kashino, K. Nakadai, T. Kinoshita and H. Tanaka, “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.

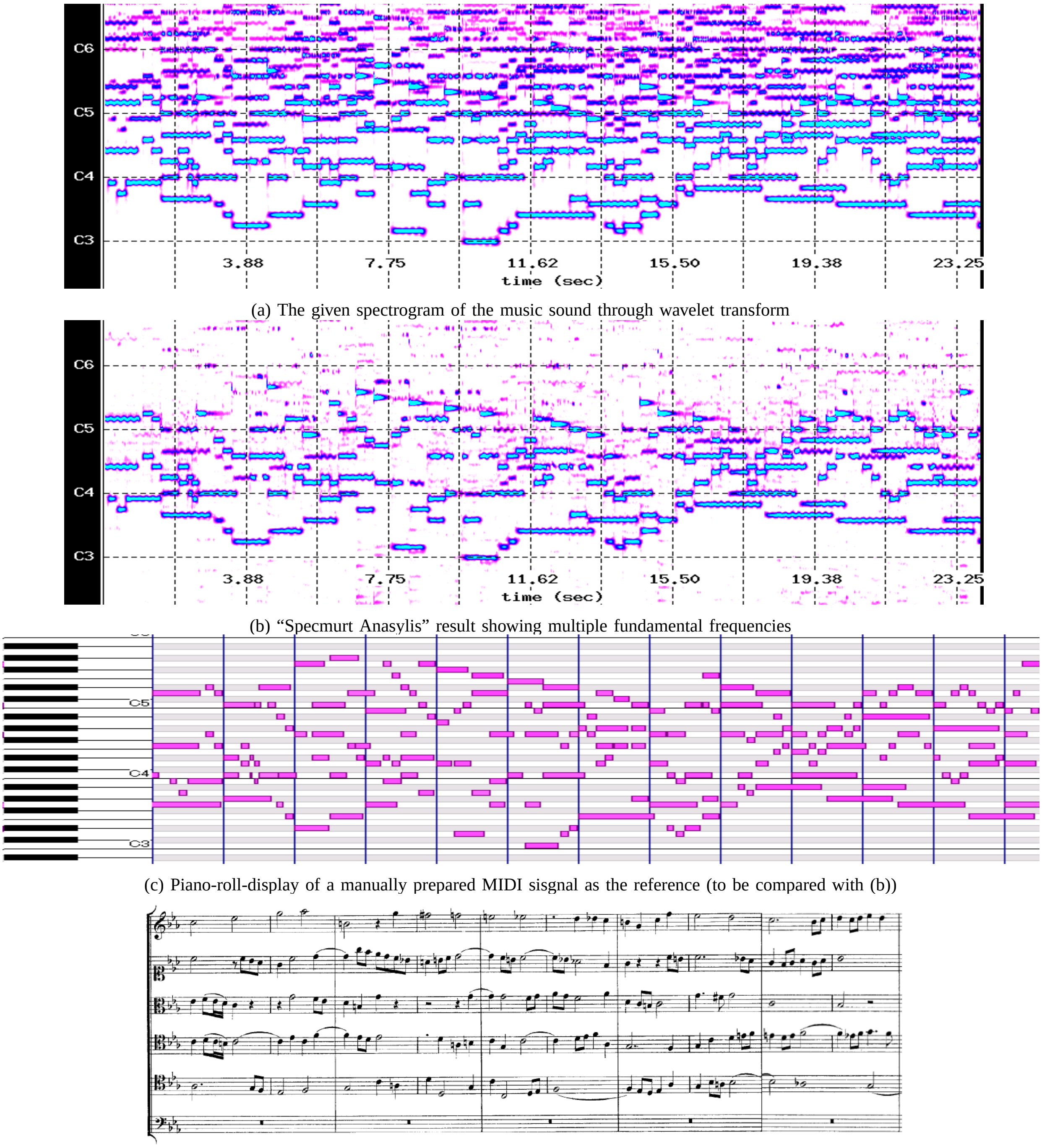
[9] S. Godsill, and M. Davy, “Baysian Harmonic Models for Musical Pitch Estimation and Analysis,” Proc. ICASSP2002, Vol. 2, pp. 1769–1772, 2002.

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

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

[12] M. Goto, “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.

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



(a) The given spectrogram of the music sound through wavelet transform

(b) “Specmurt Anasylis” result showing multiple fundamental frequencies

(c) Piano-roll-display of a manually prepared MIDI sisgnal as the reference (to be compared with (b))

(d) Roughly corresponding score, bars 19–23

Fig. 6. A result of “specmurt anasylis” applied to a real orchestral sound of “J. S. Bach: Ricercare à 6 from ‘Musikalisches Opfer,’ BWV 1079,” excerpted from the RWC music database[15]. Time axes are roughly aligned to each other.

[14] H. Kameoka, T. Nishimoto and S. Sagayama, “Separation of Harmonic Structures Based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds,” Proc. ICASSP2004, in CD-ROM, May 2004.

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

[16] H. Takeda, T. Nishimoto and S. Sagayama: “Automatic Rhythm Transcription from Multiphonic MIDI Signals,” Proc. 4th International Conference on Music Information Retrieval (ISMIR) (Baltimore, USA), Proc. ISMIR 2003, pp.263-264, Oct. 2003.