Audio Engineering Society

Convention Paper

Presented at the 128th Convention
2010 May 22–25 London, UK

Harmonic components extraction in recorded piano tones

Carmine Emanuele Cella[1]

[1] *Università degli studi di Bologna - Via Zamboni 38, 40126 Bologna, Italia*

Correspondence should be addressed to Carmine Emanuele Cella (carmine.emanuele.cella@gmail.com)

ABSTRACT

It is sometimes desirable, in the purpose of analyzing recorded piano tones, to remove from the original signal the noisy components generated by the hammer strike and by other elements involved in the piano action. In this article we propose an efficient method to achieve such result, based on adaptive filtering and automatic estimation of fundamental frequency and inharmonicity; the final method, applied on a recorded piano tone, produces two separate signals containing respectively the hammer knock and the harmonic components. Some sound examples to listen for evaluation are available on the web as specified in the paper.

1. INTRODUCTION

1.1. Motivations

During the development of a physical model of the piano, we needed to separate the hammer noise from the harmonic components in recorded piano tones for analysis purposes. After some research in literature we found, among the others, a method based on the SMS framework (sinusoidal modeling synthesis) described in [11]. Briefly, the authors propose a method which separates harmonic sounds by means of linear models for the overtone series. An iterative algorithm is used to estimate time-varying sinusoidal parameters, after a preliminary processing by a multipitch estimator that finds the number of concurrent sounds and their frequency components. While providing good results, we noticed some *time smearing* in the extracted noisy part (probably due to the FFT processing) that weakened the transients of the hammer. Since we were mainly interested in the noisy part and not in the harmonic one, we developed a method that processes sounds in time domain while retrieving information in frequency domain; this is achieved by means of a bank of peaking filters driven by some information coming from spectral analysis. A similar approach to the one described below, can be found in [6]: in that case the author used a bank of comb filters, but didn't adapt filtering parameters during time.

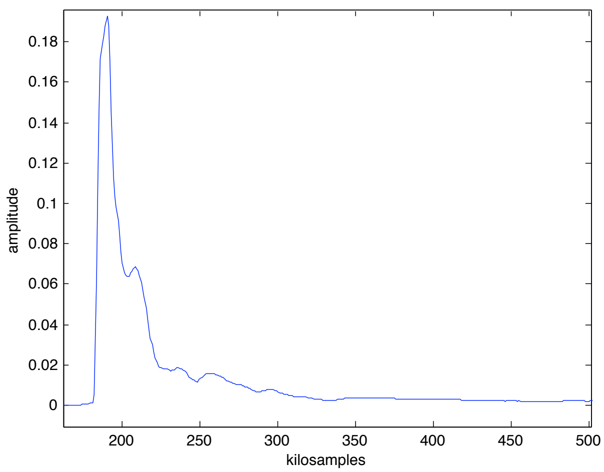

Fig. 1: The typical compound decay of piano tones.

1.2. Nature of piano tones

It's well known that piano tones have a complex nature [4]: the sound production can be described by the laws of transversely vibrating strings, but the non-linear interaction between the hammer and the string makes this description more articulated. Piano tones are basically harmonic signals that vary during time, following compound decay curves made of two major slopes in an exponential manner (usually called *double decay*). The reasons for this compound decay are mainly two: the polarization of string vibrations and the use of two, three or more mistuned strings for a single note (figure 1).

At the same time, piano tones own a certain degree of *inharmonicity* which leads to significative changes in the frequencies of real partials in respect to the harmonic series.

It is possible to calculate the frequency of the $n-th$ transverse mode of a vibrating string with the following formula [1]:

$$f_n = nf_1 = \frac{n}{2l}\sqrt{\frac{\tau}{\rho A}} \qquad n = 1, 2, \ldots \tag{1}$$

being τ the tension of the string, l it's length and n the number of the mode; A and ρ are respectively the string cross sectional area and the string density. Real strings, however, have a significant bending stiffness that creates a non-linear effect known as inharmonicity; this changes the frequency of each mode by a given amount described as:

$$f_n = nf_1\sqrt{1 + Bn^2} \qquad n = 1, 2, \ldots \tag{2}$$

where B is a factor determined by the physical properties of the string, such as material and dimensions, and is related to Young's modulus (see [1]).

During a typical hammer-string interaction, the hammer is thrown into the string thus generating pulses that spread over the string and reflect from the ends. Reflected pulses change the acceleration of the hammer and cause secondary pulses that may temporarily detouch the hammer from the string, until there are no more interactions and the string vibrates freely. Total duration for the interaction ranges typically from 0.5 ms to 5 ms. Stulov studied the piano hammer in details [9], [10] and showed that it is both non-linear and hysteretic; his model gives the force in function of compression:

$$F(t) = k\left[x(t)^p - \epsilon\int_0^t x(t-\zeta)^p e^{(\zeta/\tau_0)}d\zeta\right] \tag{3}$$

where, p is the nonlinearity constant (p = 1 means linear), x is the felt compression, τ_0 is a time constant related to hysteresis and k and ϵ are constants.

From a perceptual stand point, the transitory part of a piano tones (first few milliseconds) is really relevant; this part is mainly made of the noise generated by various pieces of the piano action, such as keys reaching the end of their travel and creating a vibration radiated throught the soundboard (usually called *thump* or *knock* noise). Spectrally, this noise has dominant peak at around 90 Hz, because of the frequency of the fundamental soundboard mode.

2. REMOVING HARMONIC CONTENT

In the previous section we showed that piano tones, while rich in inharmonicity and other non-linear effects, have a strong harmonic structure. For this reason, it is possible to achieve separation between noisy and harmonic components in piano tones by removing the harmonic part through a bank of narrow-band peaking filters tuned *around* the harmonics. Using a bank of peaking filters whose frequencies are set on harmonic frequencies and whose

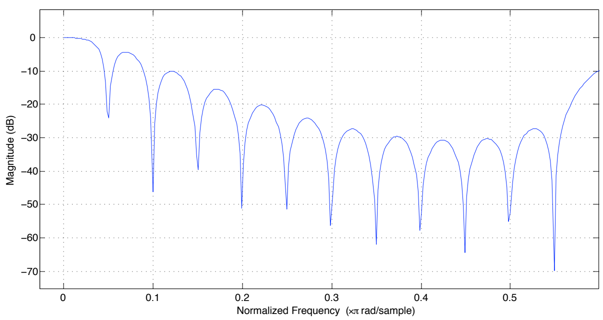

Fig. 2: Example response of the bank of filters.

gains are negative, it is possible to *equalize* the original piano tone in order to remove almost completely the harmonic content. Major problems in such approach raise in estimating the gains and the real frequencies for the filters.

2.1. Estimating the spectral shape

Piano tones have a roughly exponential spectral rolloff, determined by the initial amplitude of each mode. For this reason the gains of the filters need to compensate in some way original spectral shape in order to remove correctly the harmonic content. The right gains have been therefore computed using spectral shape detection based on the cepstrum technique. We recall that the cepstrum is calculated from the discrete Fourier transform by taking the inverse transform of the magnitude of it's logarithm; the spectral shape is then computed by applying a lowpass window to the cepstrum and by taking again the Fourier transform [12]; briefly, given the signal S:

$$E = FFT(W_{LP}(FFT^{-1}((log(|FFT(S)|))). \quad (4)$$

Once the spectral envelope is computed it is possible to evaluate approximate gains for each filter in the bank. Figure 2 illustrates an example case, for fundamental frequency of 1000 Hz.

2.2. Estimating inharmonicity

To compute the frequencies for the filters we used equation 2, where f_1 is the fundamental frequency and B is the inharmonicity factor. Having a correct B value is really critical in order to compute the real frequencies of the spectrum [3]; for this reason an estimation of the inharmonicity in the signal is performed evaluating the energy divergence of the spectral components from the multiple of the fundamental [7]:

$$I = \frac{2}{f_0} \frac{\sum_h |f(h) - hf_0| a^2(h)}{\sum_h a^2(h)}. \quad (5)$$

2.3. Estimating the fundamental frequency

In real piano tones the fundamental frequency is not completely stable and tends to oscillate because of beatings. It is obvious that if the fundamental frequency changes the computed frequencies for the filters will be wrong, leading to incorrect filtering. For this reason, a frequency estimation algorithm has been added to the proposed method in order to adapt the filtering parameters during time. After some research we choosed one algorithm between the others [2]: the *weighted auto-correlation function (WACF).*

WACF algorithm has been proposed in [5] to overcome some problems presented from the normal auto-correlation function (ACF) defined by the following equation:

$$ACF(\tau) = \phi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n+\tau). \quad (6)$$

This function measures to which extent a signal is *similar* to a delayed version of itself of time τ: since a periodic signal highly correlates with itself when the time τ is a multiple of the fundamental period, a peak in the function will be found corresponding to a period.

A complimentary way to measure the correlation of a signal with itself is described by the average magnitude difference function (AMDF):

$$AMDF(\tau) = \sigma(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} |x(n) - x(n+\tau)| \quad (7)$$

where a valley will be found corresponding to a period. In [5] Kobayashi and Shimamura noted that these functions have independent statistical behavior and can be combined to achieve more robust pitch estimation, proved to be better than ACF or AMDF alone:

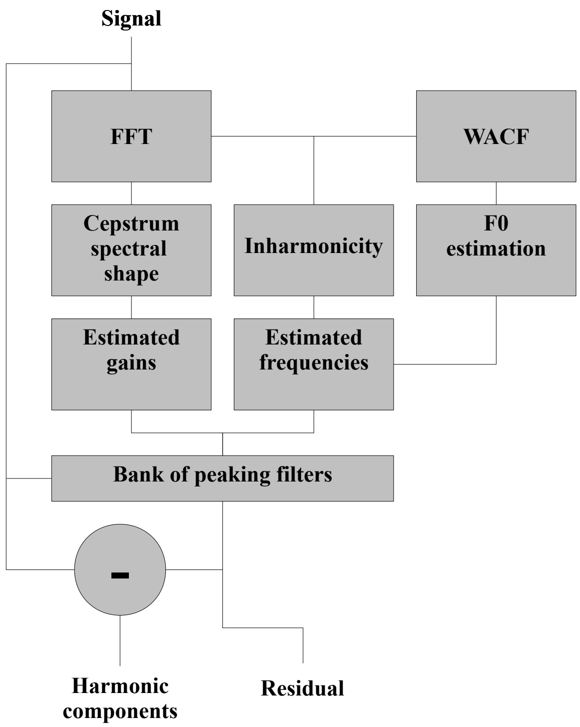

Fig. 3: The complete separation method.

$$f(\tau) = \frac{ACF(\tau)}{AMDF(\tau) + k} = \frac{\phi(\tau)}{\sigma(\tau) + k} \quad (8)$$

where k is a constant equal to 1.

Once the gains and the frequencies are correctly computed, it is also important to determine the right selectivity for the filters, ie. the q parameter. We found experimentally that this parameter depends on the fundamental frequency, roughly following three major keyboard splits:

	q	**f0**
narrow	70 to 350	A0 - A3
very narrow	351 to 500	Bb3 - A6
super narrow	> 500	$> A6$

2.4. The final method

The complete separation method, implemented in Matlab and in C++ as a command line tool, is illustrated in figure 3.

In summary, the Fourier transform is computed on the signal and then is fed into three sub-algorithms:

1. *spectral shape*: the cepstrum is computed as defined in section 2.1 and then the spectral shape is estimated;
2. *inharmonicity*: the B coefficient of equation 2 is computed as described in section 2.2
3. *F0 estimation*: by using the WACF algorithm the estimation of F0 is performed.

The information coming from the three sub-algorithms is then used to estimate real gains and frequencies for the bank of filters. The *residual* part, ie. the knock sound is obtained as result of the filtering process, while the *harmonic components* are obtained by subtracting the residual part from the original signal; the filters have been implemented as biquad sections using direct form I.

3. RESULTS AND CONCLUSIONS

Extensive testing has been done on a database of 88 real recorded notes from a Steinway & Sons grand piano. Each note has been recorded at three different levels of dynamics (*piano, mezzoforte, forte*) for a total of 264 samples. All the samples have been processed by the C++ tool, then visually and acoustically inspected. The tool outputs two different signals, containing respectively the harmonic and the noisy components. The method provided satisfying results for most of the samples, especially for highest tones (higher than A4), and the extracted hammer noise preserved its *sharpness*.

Figure 4 shows a comparison between the original sample for E5 and the residual, while figure 5 shows the same comparison for A6 sharp: in the upper spectrogram of both plots, almost all the harmonic components have been removed. For lowest tones (ie. from A0 to A2), however, the analysis revealed really small residual signals, because of the relatively small importance of the knock sound in that frequency range[1]. More samples processed by the proposed method can be found online at `http://samples.lodenstein.com` together with a version of the command line tool compiled for different platforms.

[1]For lowest tones, in fact, the hammer mass is relatively tiny compared to the string mass.

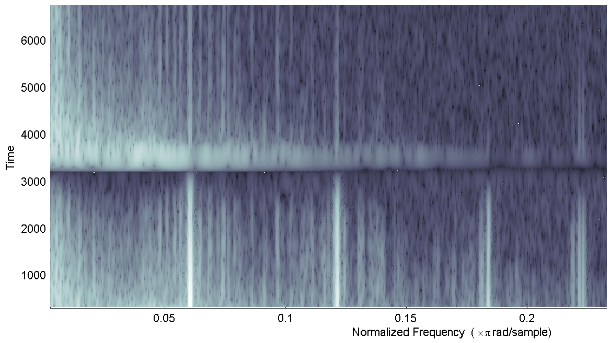

Fig. 4: Comparison between original (bottom) and residual (top) sounds for E5.

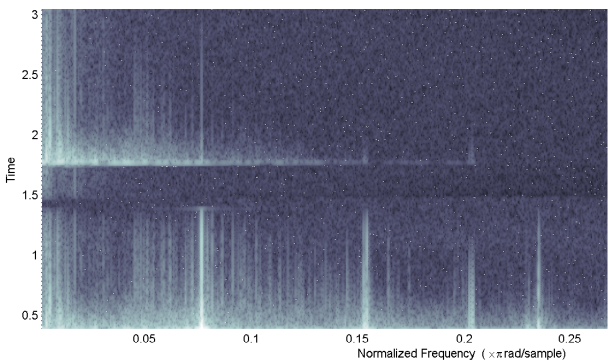

Fig. 5: Comparison between original (bottom) and residual (top) sounds for A6 sharp.

Future improvements may involve the use of different algorithms for spectral shape estimation, such as *discrete cepstrum* or *true env* [8], and for fundamental frequency estimation [2]. In conclusion, we can say that the proposed method is a reliable and efficient way to automatically separate noisy from harmonic components in piano tones.

4. REFERENCES

[1] A. Askenfelt. Five lectures on the acoustics of the piano. In *WWW edition - http://www.speech.kth.se/music/5_lectures/contents.html*, 1988 - 2000.

[2] P. de la Cuadra, A. Master, and C. Sapp. Efficient pitch detection techniques for interactive music. In *International Computer Music Conference (ICMC) - Proceedings*, 2001.

[3] H. Järveläinen, V. Välimäki, and M. Karjalainen. Audibility of inharmonicity in string instrument sounds, and implications to digital sound synthesis. In *International Computer Music Conference (ICMC) - Proceedings*, 1999.

[4] M. Keane. Understanding the complex nature of piano tone. Acoustics research centre, Auckland, 2004.

[5] H. Kobayashi and T. Shimamura. A weighted autocorrelation method for pitch extraction of noisy speech. In *Acoustics, Speech, and Signal Processing (ICASSP) - Proceedings*, 2000.

[6] O. Moravec. Separation and analysis of the musical signal noise components. 2001.

[7] G. Peeters. A large set of audio features for sound description (similarity and classification) in the cuidado project. In *CUIDADO I.S.T. Project Report*, 2004.

[8] A. Roebel and X. Rodet. Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation. In *Int. Conf. on Digital Audio Effects (DAFx) - Proceedings*, 2005.

[9] A. Stulov. Hysteretic model of the grand piano hammer felt. In *J. Acoust. Soc. (4), 2577-2585*, 1995, 97.

[10] A. Stulov. Piano hammer: Theory and experiment. In *International Symposium on Musical Acoustics (ISMA) - Proceedings, Vol 1. - 215-220*, 2001.

[11] T. Virtanen and A. Klapuri. Separation of harmonic sounds using linear models for the overtone series. In *Acoustics, Speech, and Signal Processing (ICASSP) - Proceedings*, 2002.

[12] U. Zoelzer. Digital audio effects. Wiley and sons, 2002.

[2] The authors are currently developing an algorithm called *harmonic search* based on the comparison between the highest peaks in the spectrum in order to find best candidates that fulfills some given criteria. While providing promising results, the final version of this algorithm was not ready at the time of writing this paper so has not been included in the research.