

Hybrid sinusoidal modeling of music with near transparent audio quality

Maciej Bartkowiak, Łukasz Januszkiewicz

Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznan, Poland
mbartkow@multimedia.edu.pl

Abstract— It is often believed that sinusoidal as well as sinusoidal plus noise modeling is not capable of delivering high audio quality for complex signals such as wideband music. We identify the key sources of modeling artifacts in sinusoidal modeling systems and demonstrate a hybrid system that offers near transparent quality of reconstructed audio, thanks to application of a dedicated transient model, an accurate parameter estimation method, an advanced tracking algorithm and a warped-frequency spectral model of noise.

Keywords- audio analysis/synthesis; sinusoidal model; estimation; transient modeling; noise modeling

I. INTRODUCTION

Sinusoidal model (SM) and hybrid sinusoidal + noise model (HSNM) are both well-established frameworks for signal analysis, transformation and synthesis, as well as enhancement, source separation, recognition, and data compression [1]. Since its early introduction in late 1980s, this family of techniques has been applied mostly for representing speech and single instrumental sounds [1,2,3]. Still, there is a common belief that such representation does not offer a high quality reconstructed audio for wideband signals, especially in challenging cases, like e.g. complex music with many sound sources of wide dynamics and spectra. In this paper we discuss and show a solution to several limitations of HSNM. We also present an advanced modeling system that offers a near transparent audio quality, i.e. the reconstructed signal is in most cases perceptually undistinguishable from the original audio, while a compact and meaningful parametric representation is achieved, enabling efficient implementations of auditory scene analysis, transformations, and data compression.

Generally speaking, SM consists of several stages, including spectral analysis, detection of spectral peaks identifying sinusoidal components, parameter estimation of sinusoids, and tracking of those parameters across consecutive audio frames. In this approach, all signal components are represented by modulated sinusoids, although it may be inefficient for signals with a significant amount of noise. Hybrid sinusoidal modeling addresses this problem by introducing additional modeling components. However, it requires a separation of the original signal into sinusoidal part and non-sinusoidal part, which may be particularly difficult for complex audio. In typical implementations, only certain peaks

of short time spectrum are identified as sinusoids (the deterministic component), tracked, and synthesized. The residual from the sinusoidal part represents the remaining spectral energy. This residual (noise component) may be modeled as an auto-regressive random process characterized by a time-variable spectral density function and a temporal magnitude envelope [3,4],

$$\hat{x}(t) = \underbrace{\sum_{k=1}^K A_k(t) \sin\left(\varphi_k + 2\pi \int_0^t f_k(\tau) d\tau\right)}_{\text{deterministic component}} + \underbrace{h_n(t) * \xi(t)}_{\text{noise component}}. \quad (1)$$

II. ARTIFACT SOURCES IN HSNM

Sinusoidal analysis aimed at detection of spectral peaks representing all important tonal components is usually implemented on a frame basis, as a short-time Fourier transform (STFT) followed by picking salient peaks of magnitude spectrum. STFT-based sinusoidal analysis is always a compromised solution, trading off the accuracy of representing modulated partials for the ability to capture all low-frequency partials, which is more important since they usually describe fundamental harmonics of many musical sounds and exhibit high energy. The problems with STFT are its fixed spectro-temporal resolution as well as the underlying assumption on local stationarity. High spectral resolution required for proper analysis of low pitched sounds (sometimes below 50Hz) enforces the use of long analysis windows (100-200ms, i.e. 2^{12} - 2^{13} samples if $f_s = 44.1$ kHz) in order to reliably resolve individual harmonic partials. Higher frequency components in wideband audio often exhibit deep frequency and amplitude modulations, and they must not be considered stationary within a time window of such length. On the other hand, relying on STFT with long and strongly overlapping analysis windows usually yields significant pre-echo artifacts when analyzing sounds with transients.

Transients are relatively sparse, but very important elements of sounds characterized by sudden increase of energy. Many forms of spectral processing of audio have insufficient temporal resolution that results in temporal smearing of transients, which is easily detectable and usually annoying for the listener. Since there is no way to effectively estimate transients with highly overlapping STFT frames, it may be concluded, that a separate model of transients [6] with transient

removal before sinusoidal analysis, as well as transient-aware multi-resolution sinusoidal analysis, are both necessary for high quality representation.

In a typical HSNM system, only spectral peaks representing actual sinusoids should be selected for the tonal part of the model, and their parameters should be tracked in order to establish sinusoidal trajectories. In practice, discerning between sinusoidal and stochastic spectral peaks is a challenging problem. First of all, the bulk of spectral components observed in natural audio is neither purely sinusoidal nor purely random. Several techniques for classification of spectral peaks have been proposed [7,8], but the general experience is that applying such selection is always prone to misclassification and audible modeling errors. In our experience, the best verification of whether given spectral peak is a sinusoidal one, is if it yields a reliable knot of a sinusoidal trajectory as a result of tracking. Therefore, application of any peak classification criteria should be very conservative in order to avoid rejecting weak sinusoidal partials which may be obscured by noise.

Accurate partial tracking is probably the most challenging problem in HSNM, because the goal is not well defined and it depends on particular application. For example, too fragmented trajectories resulting from too conservative connection rules yield a model that is inefficient in terms of data compression. Conversely, long and continuous trajectories obtained by excessive linking of partial data representing actually different sources may result in significant errors in source separation. In our experience, simple tracking algorithms [2,3] are inappropriate for modeling of wideband music because of not taking into account the wider temporal context of established connections and because of too simplified connection criteria, depending mostly on absolute frequency difference and not reflecting deeper modulations observed in upper harmonics. Trajectories obtained from a simple tracking algorithm are usually fragmented and chaotic. A signal synthesized from such a model is inferior in quality due to many audible discontinuities of partials which cannot be easily masked by applying a smooth fade-in and fade-out to segments at trajectory endpoints, since such amplitude modulation introduces a significant spectral distortion. For the sake of preserving the continuity as much as possible, tracking based on various forms of adaptive prediction is preferable in high quality HSNM.

Handling the non-sinusoidal component by a separate model requires obtaining the residual of SM as clean and free from unwanted sinusoids as possible, because otherwise the residual spectral model tends to compensate for their energy, and the amount of noise becomes overestimated. The residual may be derived as a time-domain difference between the original signal and the synthesized sinusoids [2], or through spectral subtraction [3]. The first option requires an accurate estimation of partial parameters, as well as phase-coherent synthesis. The latter option is more tolerant to estimation errors, but it usually yields the power density spectrum being underestimated.

Spectral modeling of the residual, interpreted as a random noise, is often implemented in a form of auto-regressive modeling, or linear prediction (LP). Unfortunately, this popular

technique [1,3,4] is not well suited for modeling colored noise components in wideband music, because of its frequency resolution being uniform in a linear scale which does not match the non-uniform resolution of human ear. Hence, an LP model of reasonable order is very inaccurate in the low frequency range while it is unnecessarily accurate in high frequencies.

III. SINMOD TOOLBOX

A hybrid sinusoidal modeling system has been developed for dealing with wideband complex music signals in the HSNM framework. The software implementation in a form of a Matlab toolbox is freely available for non-commercial applications at http://www.multimedia.edu.pl/audio_research/.

It has been verified through a number of blind listening experiments, that this system offers a near transparent quality, i.e. the reconstructed audio is perceptually nearly undistinguishable from the original music recording. The key elements that contribute to this high fidelity are:

- a dedicated transient model, with transients re-synthesized and removed from the signal prior to sinusoidal analysis,
- multi-resolution sinusoidal analysis for detection of both low frequency dense partials and deeply modulated higher frequency partials,
- adaptive prediction-based partial tracking for creating long, continuous and meaningful trajectories,
- post-processing of sinusoidal trajectories to cope with overestimation and fragmentation of trajectories,
- accurate sinusoidal parameter re-estimation once tracks are established, enabling accurate and phase-coherent synthesis,
- a noise model with frequency resolution corresponding to the resolution of human ear.

These elements will be discussed in the remaining part of this paper.

A. Modeling of transients

Before transient modeling, a reliable detection is to be performed. Popular transient detection techniques are based on thresholding of certain audio features, like local energy or spectral flux. In the HSNM system proposed here, a complex spectral domain prediction (2) is employed [9] for detecting sudden changes of signal short-time amplitude and phase spectrum, usually associated with discontinuities, note onsets, or short bursts of energy accompanying transients,

$$\eta(m) = \sum_{k=1}^K |X_k(m) - \hat{X}_k(m)|, \quad (2)$$

where $\hat{X}_k(m) = |X_k(m-1)| \exp[j2\varphi_k(m-1) - j\varphi_k(m-2)]$ is a complex-valued prediction of a DFT co-efficient $X_k(m)$ based on two previous frames, $m-1$, and $m-2$.

The detection function $\eta(m)$ as proposed in [9] is correlated with signal magnitude, therefore an adaptation to local signal

dynamics is necessary for reliable detection. A decision process with an adaptive hysteresis is used here. The lower and upper thresholds are dependent on local mean and median values of $\eta(m)$. Furthermore, in order to avoid false alarms on pure noise, transient detection is enabled only when signal amplitude exhibits a significant local peak.

For transient modeling, a simple model of damped sinusoids sharing a common amplitude envelope is adopted from [10]. In the first step, a parameterized envelope model (so called Meixner function) is fitted to the magnitude of the signal within a short rectangular window. Subsequently, a set of sinusoidal modulating components is iteratively detected based on FFT analysis of the original signal windowed by the envelope determined in the first step. Finally, the phase of each sinusoid is estimated using a least-squares fit.

The procedure results in a set of data consisting of three envelope parameters, identifying the position, attack time and decay time of each transient, as well as frequencies, amplitudes and phase of each of the modulating sinusoids. The signal synthesized from these parameters matches the original waveform and may be subtracted in time domain resulting in a conditioned signal that is better suited for sinusoidal analysis (cf fig. 1).

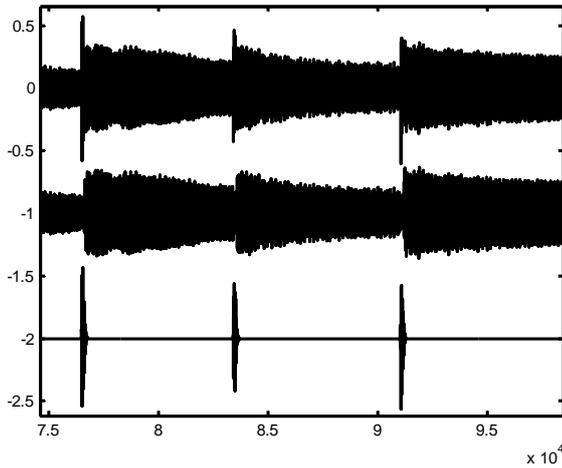


Figure 1. Example of transient synthesis from parameters (bottom) and the signal after transient subtraction (middle) from the original signal (top).

B. Multi-resolution sinusoidal analysis

Accurate detection of spectral peaks representing partials in the dense range of very low frequencies as well as detecting deeply modulated partials in the more sparse range of upper frequencies calls for multi-resolution spectral analysis. The proposed HSNM system employs a traditional structure of subband decomposition followed by FFT transform of different resolutions suited to particular signal properties in each subband. For practical reasons, the number of subbands is limited to three. The configuration (splitting frequencies and transform block lengths, N) is experimentally optimized by calculating the modeling error for a range of signals with all reasonable combinations of settings. Listening tests indicate, that the best configuration found in this way (table I) also offers a best subjective quality of modeling. In all resolutions,

consecutive analysis frames are strongly overlapping and advanced by 6 or 12ms.

TABLE I. MULTI-RESOLUTION ANALYSIS SETUP

Subband	Frequency range	Subsampling	FFT length, N
1	20Hz – 310Hz	64:1	16384 (256)
2	310Hz – 2480Hz	8:1	2048 (256)
3	2480Hz – 20kHz	–	1024

Transient detection and removal before sinusoidal analysis effectively reduces pre-echo in case of impulse-like transients, however there is still a possibility that high window overlapping yields pre-echo in case of step-like transients. Therefore, a special pre-processing is applied for analysis windows marked as containing a transient in their second half. In such a case, the sequence of samples starting from detected transient position is replaced by a predictor output based on previous samples (fig. 2). A high-order (e.g. $p=500$) autoregressive (LP) model is trained on data samples preceding the transient, and a sequence of zeros is passed to the input of the LP predictor while preserving its inner state after processing those original samples. The prediction signal partially replaces the original signal. This allows to avoid detection of new sinusoidal partials related to transient in frames preceding the actual transient position.

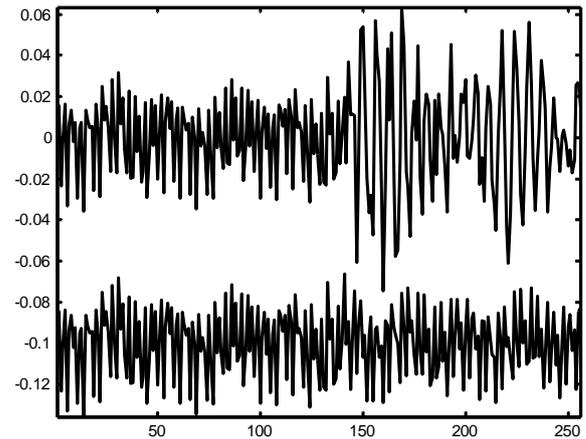


Figure 2. Example of step-like transient removal. In the case of a transient located in the second half of the analysis window, the original signal (top) is replaced by the output of a high order predictor (bottom).

Sinusoidal partials are detected by applying the standard peak picking procedure to the magnitude spectrum in each frame. Optional peak selection may be performed in order to avoid estimation and further tracking of partials which are inaudible. For this purpose, all peaks falling below the frequency dependent absolute threshold of hearing are rejected. Furthermore, a small clearance zone (e.g. 10Hz) is defined around each detected peak. Peaks of magnitude lower than 10dB w.r.t the maximum peak within this zone are rejected as well.

Estimation of partial frequencies, amplitudes and phase is based on the ML method with quadratically interpolated Fourier transform and takes into account the shape distortion of

spectral main lobes related to frequency and amplitude modulations [11]. However, the proposed HSNM system is not fixed to any particular estimation method, and other methods may be used as well.

C. Adaptive tracking

The tracking algorithm is a result of an extensive development. It applies a carefully chosen set of criteria for finding track continuations in a collection of spectral data. The most important technique employed here is an adaptive prediction which is much more successful in tracking of modulated sounds than any simple technique taking into account only the absolute frequency difference of partials. An LP predictor is capable of learning the character of typical vibrato and tremolo from the beginning of the note and accurately predicts its further evolution [12]. Its application is motivated by the observation that pitch and intensity variations in many natural sounds are related to the motion (rocking or swinging) of player's hand which in turn may be characterized by certain mechanical resonant modes.

For each trajectory defined by a sequence of parameters $\{f_i, A_i\}$, the continuation is calculated from its existing evolution with a standard LP prediction equation,

$$\hat{f}_m = \sum_{i=1}^P a_i^{(f)} f_{m-i} \quad \text{and} \quad \hat{A}_m = \sum_{i=1}^P a_i^{(A)} A_{m-i}. \quad (3)$$

For all data points available in the current frame, a degree of frequency and amplitude matching λ is calculated by

$$\lambda_f \{\hat{f}_m, f_m\} = \min \left\{ 1 - \frac{|\hat{f}_m - f_m|}{\Delta_{\max} f_m}, 0 \right\}, \quad (4)$$

where $\Delta_{\max} f_m = \max \{ \delta_f f_m, \Delta_f \}$ [Hz], and

$$\lambda_A (\hat{A}_m, A_m) = \min \left\{ 1 - \frac{|\hat{A}_m - A_m|}{\Delta_{\max} (\hat{A}_m, A_m)}, 0 \right\} \quad (5)$$

where $\Delta_{\max} (\hat{A}_m, A_m) = \begin{cases} \Delta^+ A & \hat{A}_m \geq A_m \\ \Delta^- A & \hat{A}_m < A_m \end{cases}$ [dB].

The above measures are normalized in the range of $\langle 0, 1 \rangle$, and related to predefined thresholds ($\Delta f, \delta f, \Delta^+ A, \Delta^- A$) that allow to control the sensitivity of the algorithm.

Note, that for the maximum change of frequency $\Delta_{\max} f$, both absolute difference limit (Δf) and relative difference limit (δf) is considered (set approximately to 30Hz and 3%, respectively). This is a crucial modification w.r.t. the original algorithm [2], and allows to properly cope with frequency modulation depth increasing for high-order partials of a sound spectrum, while taking into account the typical accuracy limitations of frequency estimation which is a part of sinusoidal analysis. On the other hand, for the maximum amplitude change $\Delta_{\max} A$, separate limits are defined for amplitude increase ($\Delta^+ A$) and decrease ($\Delta^- A$), typically in the range of 6 to 20 dB. The joint degree of matching is calculated as $\lambda = (\lambda_f \lambda_A)^{1/2}$.

Connections are made according to a greedy rule, i.e. best matching pairs are connected in the first order, and a connection is forbidden for $\lambda=0$.

Other track continuation methods may be used optionally, such as the first order, non-adaptive prediction, where $a_i=0$ for $i>1$, or a linear trend-based prediction in log scale of frequency and amplitude. In such a case, the algorithm resorts to alternative criterion when the basic criterion does not find a matching data point. The last resort is generating zombie points that help to bridge connections over a number of frames with missing data. A sequence of limited number (e.g. 2 or 3) of successive zombie points is allowed in each trajectory by simply using the predictor outputs for f_m , and A_m , respectively.

D. Merging of trajectories

The tracking algorithm creates trajectories progressively, from previous frame to the current frame, in the direction of time. This strategy may result in missing connections, due to bad initialization of the predictor (3). Furthermore, in certain conditions, a sequence of alternating values representing a modulated partial yields a creation of several parallel trajectories with zombie points instead of one evolving according to the modulations. An additional post-processing of trajectories is aimed at increasing the continuity by merging fragmented trajectories as well as absorbing weak trajectories by a close strong neighboring one.

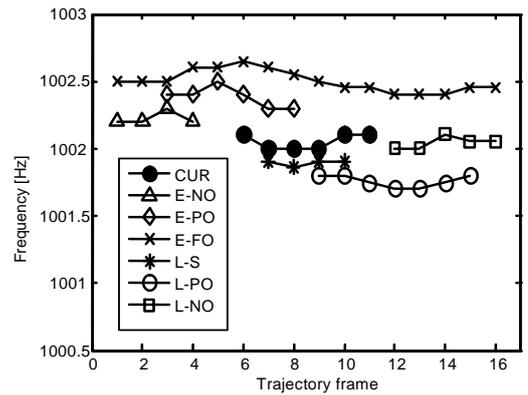


Figure 3. The classes of neighboring trajectories.

The iterative trajectory merging algorithm analyses all frames in a sequence. A set of trajectories which end in current frame is determined and sorted according to the energy of corresponding sinusoids. For every currently considered trajectory (CUR) a list of merging candidates is created. All trajectories within a small time and frequency neighborhood of CUR are assigned a specific class. Six classes are defined (cf Figure 3.): earlier–non–overlapping (E-NO), earlier–partially–overlapping (E-PO), earlier–fully–overlapping (E-FO), later–shorter (L-S), later–partially–overlapping (L-PO) and later–non–overlapping (L-NO). The best possible candidate for merging is determined in next step. All L-S candidates are discarded at the beginning, as they are redundant in current iteration. The choice between the rest of candidates is based on a degree of matching, which is essentially the same as defined in (4-5), albeit with more conservative limits of Δf and δf (10Hz and 1%, respectively). For non–overlapping

candidates, an LP based extrapolation of trajectories is calculated for a number of frames in order to determine the degree of matching on a longer distance.

The actual process of merging varies depending on whether the accepted candidate is overlapping or non-overlapping. For overlapping cases, all overlapping knots are to be combined and their parameters need to be recalculated. The new values of amplitudes and frequencies are determined as

$$\hat{A} = |\hat{X}|, \quad \hat{\phi} = \arg\{\hat{X}_k\}, \quad \hat{f} = \frac{A_k^2 f_k + A_m^2 f_m}{A_k^2 + A_m^2}, \quad (6)$$

where $\hat{X} = |X_k + X_m| \exp(j \arg\{X_k + X_m\})$, $X_k = A_k \exp(j\phi_k)$ and $X_m = A_m \exp(j\phi_m)$ are complex representations of corresponding trajectory knots. In case of non-overlapping candidate choice, missing knots are obtained by linear interpolation of amplitude and cubic spline interpolation of frequency values. An example effect of trajectory merging is shown in Figure 4. It may be noted that this process results in combining a sequence of broken segments into a long continuous trajectory.

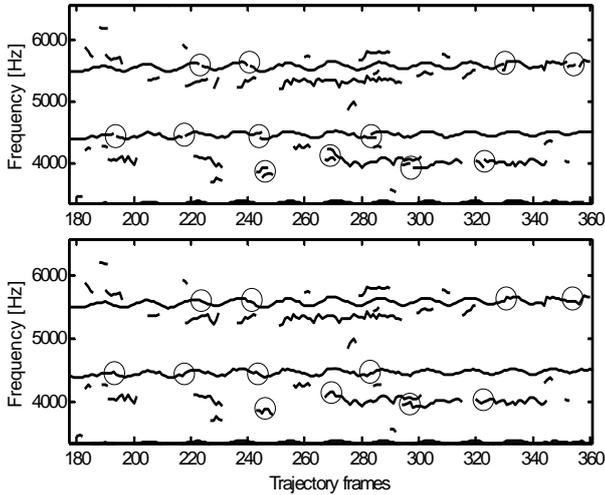


Figure 4. Trajectories before (top) and after merging (bottom). Thin-line circles indicate places where the merging was performed.

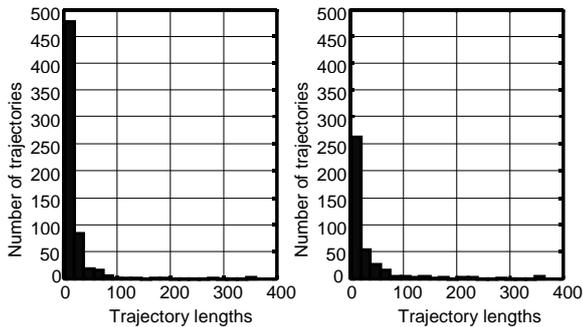


Figure 5. Example histograms of trajectory lengths before merging (left) and after merging (right).

Fig. 5 shows that after trajectory merging the number of short trajectories is significantly reduced, while the number of longer trajectories is slightly increased. The total number of trajectories after merging is usually reduced by 40%-60%. The advantage of merging is not only the reduced complexity of the model, but also reduced number of discontinuities which are responsible for potential synthesis artifacts.

E. Parameter re-estimation

The additional stage of parameter estimation aims at correcting estimation errors, potential artifacts resulting from merging, as well as estimating zombie data points. Having the trajectories already established, it is possible to estimate the amplitudes and frequencies of time-varying sinusoids yielding minimum energy of the residual. The re-estimation is performed frame by frame. In every data segment centered at current frame, a least-squares solution to a matrix equation $\underline{x}_+ = \mathbf{W} \underline{c}$ is computed, where

$$\mathbf{W} = \begin{bmatrix} A_1(1) \exp[j\phi_1(1)] & \cdots & A_K(1) \exp[j\phi_K(1)] \\ \vdots & \ddots & \vdots \\ A_1(N) \exp[j\phi_1(N)] & \cdots & A_K(N) \exp[j\phi_K(N)] \end{bmatrix} \quad (7)$$

is a matrix of interpolated trajectory samples, \underline{c} is a vector of complex correcting coefficients for all trajectories in the current frame, \underline{x}_+ is a column vector of samples of an analytic signal $x_+(t) = x(t) + j \mathcal{H}\{x(t)\}$, and $\mathcal{H}\{\}$ is the Hilbert transform.

The elements of \underline{c} obtained by solving the above equation are used to correct partial parameters by substituting

$$A_k \leftarrow A_k |c_k|, \quad \text{and} \quad \phi_k \leftarrow \phi_k + \arg\{c_k\}. \quad (8)$$

The choice of the length of segment N is quite important as it affects the accuracy of the technique in time and frequency. It is particularly essential in the low frequency range, where the segment should be long enough to accommodate many cycles of the estimated waveform. For best results, the set of trajectories is divided according to their mean frequency, and different segment lengths are used in each subset. Analysis settings shown in table 1 proved in an extensive series of experiments to deliver the most accurate results.

F. Noise modeling

The noise model is based on the frequency warped LP technique [13]. It is particularly advantageous for wideband audio, since a proper selection of the warping coefficient allows to achieve a much more accurate estimation of the residual spectrum envelope in the low frequency range than the traditional auto-regressive (LP) model (cf. fig. 6).

The residual is analyzed in consecutive overlapping frames of fixed length (e.g. 10ms). The warped linear prediction coefficients are estimated in each frame as well as the energy of the prediction error. During re-synthesis, these parameters are used to generate an appropriately shaped random signal. Additional HP filter (2nd order, $f_c=300\text{Hz}$) and an LP filter (2nd order, $f_c=12\text{kHz}$) are employed for compensating the over-estimated power density in the very low and very high frequencies, as shown in fig. 6.

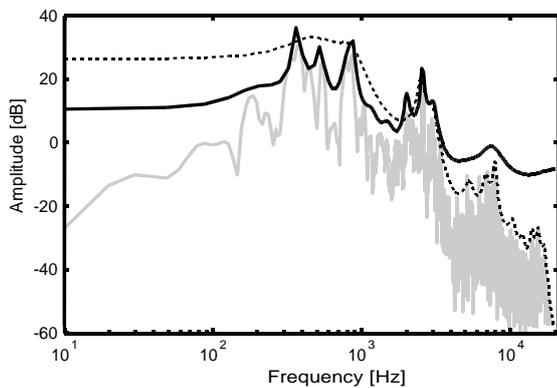


Figure 6. Spectral envelope modeling of the original signal (grey) by a standard LP model (dashed line) and the WLP model (black) of the same order (30). Note the increased accuracy in low frequency range in WLP versus an unnecessarily high accuracy in the high frequency range in LP.

G. Signal synthesis from the model

The synthesis procedure is quite straightforward and may be performed in any order. Sinusoidal partials are synthesized sample by sample from the amplitude, frequency and phase data after linearly interpolating the amplitude in the log (dB) scale, and interpolating the phase with a cubic spline polynomial [2]. For pitch or speed transformations, phase data needs to be recalculated based on the integral of instantaneous frequency. Noise and transient components are synthesized from respective data sets and mixed with the sinusoidal part.

IV. RECONSTRUCTION QUALITY

The HSNM system described in this paper has been thoroughly tested and the reconstruction quality has been assessed in many experiments. For the purpose of this paper, a blind listening test has been organized according to the MUSHRA methodology [14]. Twelve subjects participated in the test, evaluating in a continuous subjective scale the anonymous reconstructed signal against known as well as a hidden reference (the original), for a collection of music excerpts representing various modeling challenges: a solo piano, solo harpsichord, a violin+acoustic guitar, a vocal quartet, a dynamic pop and RnB music, a choral piece, and a symphonic orchestra piece. The results (cf. fig. 7) indicate, that in many cases the listeners could not reliably identify the re-synthesized signal.

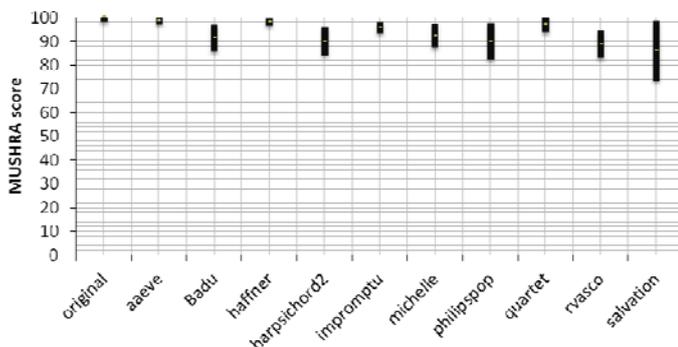


Figure 7. Blind listening test results: individual items scores are shown with 95% confidence intervals.

The reader may also individually assess the quality of reconstructed signals by visiting the project homepage at http://www.multimedia.edu.pl/audio_research/.

V. CONCLUSIONS

A hybrid sinusoidal modeling system that offers a near transparent audio quality even for complex and dynamic music has been described in the paper. This high reconstruction fidelity is achieved thanks to introduction of a dedicated transient model, as well as numerous enhancements within the traditional sinusoidal modeling scheme. The applications of this system include high quality parametric audio coding, source separation, pitch and time scale transformations, and other special effects.

ACKNOWLEDGMENT

The work has been supported by publish funds for scientific research.

REFERENCES

- [1] J. Beauchamp, "Analysis and Synthesis of Musical Instrument Sounds," in Analysis, Synthesis, and Perception of Musical Sounds, J. Beauchamp, Ed. Urbana: Springer, 2007
- [2] R. McAulay and T. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation", IEEE Trans. Acous., Speech, Sig. Proc., vol. 34, no. 4, Aug. 1986
- [3] X. Serra and J.S. Smith III, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition," Computer Music Journal, vol. 14, no. 4, 1990
- [4] M. Goodwin, "Residual modeling in music analysis-synthesis," Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP-96, 7-10 May 1996
- [5] X. Rodet, "Musical Sound Signal Analysis/Synthesis: Sinusoidal+Residual and Elementary Waveform Models," IEEE Time-Frequency and Time-Scale Workshop, Coventry, UK, 1997
- [6] R. Badeau, R. Boyer and B. David, "EDS Parametric Modeling and Tracking of Audio Signals", Int. Conf. on Digital Audio Effects, DAFx'02, September 2002
- [7] G. Peeters, X. Rodet, Signal Characterization in terms of Sinusoidal and Non-Sinusoidal Component, Proc. 1st Digital Audio Effects Conference (DAFx'98), Barcelona, 1998
- [8] G. Peeters, X. Rodet, SINOLA: A New Analysis/Synthesis Method using Spectrum Peak Shape Distortion, Phase and Reassigned Spectrum, in Proc. Int. Computer Music Conf., Beijing, October 1999
- [9] C. Duxbury, J. P. Bello, M. Davies and M. Sandler, "Complex Domain Onset Detection for Musical Signals," 6th Int. Conference on Digital Audio Effects DAFx'03, London, UK, 2003
- [10] A.C. den Brinker, E.G.P. Schuijers, and A.W.J. Oomen, "Parametric Coding for High-Quality Audio", 112th Convention of the Audio Engineering Society, Munich, 2002
- [11] M. Abe, J. O. Smith III, "Design Criteria for Simple Sinusoidal Parameter Estimation Based on Quadratic Interpolation of FFT Magnitude Peaks," Proc. 117th Convention of the Audio Engineering Society, San Francisco, Oct. 2004
- [12] M. Lagrange, S. Marchand, M. Raspaud, J. B. Rault, "Enhanced Partial Tracking Using Linear Prediction," Int. Conf. on Digital Audio Effects, DAFx'03, London, Sept. 2003
- [13] A. Härmä, et al, "Frequency-Warped Signal Processing for Audio Applications", J. Audio Eng. Soc., vol. 48, no.11, 2000
- [14] ITU-R, "Method for the subjective assessment of intermediate quality level of coding systems", ITU-R, Tech. Rep. BS. 1534-1, Rec. 2003