A ROBUST ALGORITHM FOR PARTIAL TRACKING OF MUSIC SIGNALS

Hamid Satar-Boroujeni and Bahram Shafai

Electrical and Computer Engineering Department
Northeastern University, Boston, MA, USA
{hsattar&shafai}@ece.neu.edu

ABSTRACT

In this paper we propose a novel approach for tracking of partials in music signals based on a robust Kalman filter. Our tracker is based on a regularized least-squares approach that is designed to minimize the worst-possible regularized residual norm over the class of admissible uncertainties at each iteration. We introduce a set of state-space models for our signals based on the evolution of frequency and amplitude in different classes of musical instruments. These prior models are used to estimate future values of partial tracks in successive time frames of our spectral data. Here, the parameters of evolution models are treated as bounded uncertainties and our tracker can robustly track partials in all frequency regions. Unlike the conventional Kalman tracker, performance of this tracker is not influenced by the magnified track variations in higher frequencies. This tracker promises an improved performance over conventional Kalman tracker while preserving its good properties and superiority over existing methodologies.

1. INTRODUCTION

Partial tracking has been widely used in different areas of music signal analysis where prominent features of these signals, such as pitch and frequency-amplitude of harmonics are extracted. The role of partial tracking in all these areas can be boiled down to an attempt for tracking time-varying features in separate analysis frames of a continuous-time music event. These features are captured from estimated spectrum for frames of the temporal data that are small enough to be assumed as stationary.

There are various methodologies for tracking of partials in audio signals, all of which are based on a model of pseudo-stationary sinusoidal plus noise [1]. Partial tracking was first used in analysis and synthesis of speech signals [2] and then adopted for the case of music signals [1], where it was based on a heuristic approach. In a more recent approach [3] and as an extension to [2], linear prediction was used to enhance the tracking of frequency components in music signals. In all these approaches peaks from successive frames are connected to each other based on their proximity in frequency, and the behavior of peaks' amplitude is not taken into account while performing the tracking. Another approach [4], which was inspired by a similar technique in radar tracking and also a frequency tracker for avalanche signals [5], takes the advantage of Kalman filter by constructing a state-space model for the behavior of peaks' power (i.e. amplitude in dB scale) and frequency. In this approach peaks are not matched based on how close they look like in frequency, rather they are matched based on the future behavior of a peak's frequency and power.

We proposed a partial tracking technique before [6], which was based on the conventional Kalman filter. Parameters of the evolution models for this system were estimated through a statistical analysis of a large database of musical sounds and by averaging over varying estimates. This inaccuracy in model parameters, which is unavoidable when dealing with real world models, degraded the performance of our tracker in certain situations. This sensitivity of Kalman filter to model parameters has also been studied before [7].

A feasible solution to this problem can be the use of a robust Kalman tracker which deals with model parameters as bounded uncertainties. This can be especially rewarding since we do not need to tediously estimate these parameters for different situations where they can never be accurate enough and, on the other hand, our robust tracker can perform a significantly better job in critical situations (as will be shown in the results section).

A general block diagram for the process of forming frequency and power partial tracks from the given sound wave is presented in Figure 1.

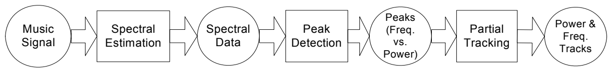

Figure 1: The process of partial tracking

This paper will proceed with a brief discussion on the problem of peak detection. In section 3 we will discuss the problem of music signal modeling and introduce a set of state-space models. The formulation of our robust tracker, which is based on the approach of [8], will be discussed in section 4. In section 5 we include some results and compare the performance of this tracker with methods using conventional Kalman tracker. A useful approximation in reducing the computational expense of our algorithm will follow at the end.

2. PEAK DETECTION

Detection of peaks in spectral representations is a very important task in the process of music signal analysis. In its ideal shape, a peak detection algorithm must be able to detect all the peaks pertaining to existing partials and rule out all those that are most likely related to noise or imperfections in estimating the spectrum. Optimum number of peaks will optimize the computational load of the tracking process. On the other hand, a large number of inaccurate peaks can result in formation of false partial tracks from randomly successive sets of spurious peaks. Based on these requirements we proposed a novel technique before [9].

Our proposed algorithm consists of two steps which are shown in Figure 2 . In the first step we use a mathematical framework to collect all the peaks that fit into the very definition of a peak as a local maximum. These are referred to as raw peaks. In the next step, statistical properties of a relative number of data points surrounding each raw peak are used to examine the concreteness of detected peak and reject any incompetent maxima.

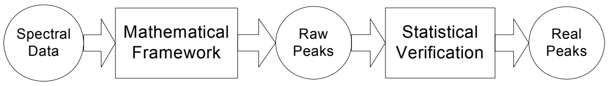

Figure 2: *Process of peak detection in two steps*

The output of the peak detector is power and frequency information of each peak which is stored in two vectors for each time frame.

3. MODELING

3.1. Time-Varying Partials

A well-known approach to modeling of music signals for the purpose of statistical analysis/synthesis assumes a model of additive sinusoidal plus residuals that can be formulated as [1]

$$y(t) = s(t) + e(t) \tag{1}$$

with

$$s(t) = \sum_{n=1}^{N} A_n(t) \cos\left(\omega_n(t) + \phi_n(t)\right) \tag{2}$$

Here, $s(t)$ reflects the pure musical part of the signal and $e(t)$ represents the additive noise, which can be modeled as a stationary autoregressive process. In the musical portion, $A_n(t)$ and $\omega_n(t)$ are representatives of time-varying amplitude and frequency of partials, and N is the number of partials. Quantity $\phi_n(t)$ represents timbral variations and performance effects. In our analysis this can be considered as a noise process.

3.2. Evolution Models

What we have as observation is discrete sets of peaks from successive time frames. $A_n(t)$ and $\omega_n(t)$ can be estimated by making connections between those peaks from adjacent frames that look like being the continuation of the same partial.

Kalman filtering takes the noisy observations and based on a model for evolution of certain states finds the optimal estimate of the process behavior. Here, the noise corrupted observations are the identified peaks and system model is a state-space model for evolution of frequency and power. This model can be represented as

$$\begin{aligned} x(k+1) &= Ax(k) + Bv(k) \\ y(k) &= Cx(k) + w(k) \end{aligned} \tag{3}$$

where

$$\begin{aligned} x(k) &= \begin{bmatrix} f(k) & p(k) & n_1(k) & \cdots & n_m(k) \end{bmatrix}^T \\ v(k) &= \begin{bmatrix} u_1(k) & \cdots & u_m(k) \end{bmatrix}^T \\ y(k) &= \begin{bmatrix} f(k) & p(k) \end{bmatrix} \end{aligned} \tag{4}$$

Here, $f(k)$ and $p(k)$ are frequency and power for a detected peak respectively. $v(k)$ and $w(k)$ are process noise and observation noise, and $n_i(k)$, $i=1,\dots,m$ are states for as many shaping filters for which the uncorrelated noise processes $u_i(k)$, $i=1,\dots,m$ are white. The matrix A is the transition matrix, B describes coupling of the process noise $v(k)$ into the system states, and C is the observation matrix. In this model, $v(k)$ and $w(k)$ are zero-mean and jointly uncorrelated Gaussian processes with covariance matrices Q and R, respectively.

For specifying matrices and number of states needed for our modeling, prior information about the power and frequency partials is needed. This can help us to specify the model by a piecewise-linear fit to $p(t)=20\log A_n(t)$ and $f(t)= \omega_n(t)/2\pi$.

Based on the overall shape of frequency and power partial in different classes of instruments, we introduced two groups of models for the purpose of Kalman tracking before [10]. For the class of instruments with nearly constant frequency and power partials, which are called the class of *Continued Energy Injection* (CEI), the sate-space model is as follows

$$\begin{aligned} f(k+1) &= f(k) + n_1(k) \\ n_1(k+1) &= a_1 n_1(k) + b_1 u_1(k) \\ p(k+1) &= p(k) + n_2(k) \\ n_2(k+1) &= a_2 n_2(k) + b_2 u_2(k) \\ x(k) &= \begin{bmatrix} f(k) & p(k) & n_1(k) & n_2(k) \end{bmatrix}^T \\ v(k) &= \begin{bmatrix} u_1(k) & u_2(k) \end{bmatrix}^T \\ y(k) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x(k) + w(k) \end{aligned} \tag{5}$$

For the instruments with constant frequency partials and linearly decaying power partials, which are in the class of *Discontinued Energy Injection* (DEI), the state-space model is

$$\begin{aligned} f(k+1) &= f(k) + n_1(k) \\ n_1(k+1) &= a_1 n_1(k) + b_1 u_1(k) \\ p(k+1) &= p(k) + v_p(k) \\ v_p(k+1) &= v_p(k) + n_2(k) \\ n_2(k+1) &= a_2 n_2(k) + b_2 u_2(k) \\ x(k) &= \begin{bmatrix} f(k) & p(k) & v_p(k) & n_1(k) & n_2(k) \end{bmatrix}^T \\ v(k) &= \begin{bmatrix} u_1(k) & u_2(k) \end{bmatrix}^T \\ y(k) &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} x(k) + w(k) \end{aligned} \tag{6}$$

In a polyphonic setting there are three possible scenarios when we do not consider non-melodic instruments such as drums. A piece of music can consist of instruments solely form the CEI, or only from DEI or a combination of both.

Depending on the polyphonic scenario, we can estimate the parameters of each model, e.g. a_1, b_1, a_2, b_2, by performing a statistical analysis on a large number of musical sounds with known identities and within the specific scenario in a forward-problem setting. The details of this procedure are presented in [10].

Based on our experience, these parameters are frequency dependant. Therefore, in each class and for different frequency bins we have different sets of parameters. Estimated parameters for the CEI class are shown in Figure 3. In section 5, where we transform our model to a form appropriate for robust Kalman tracker, these parameters are treated as bounded uncertainties.

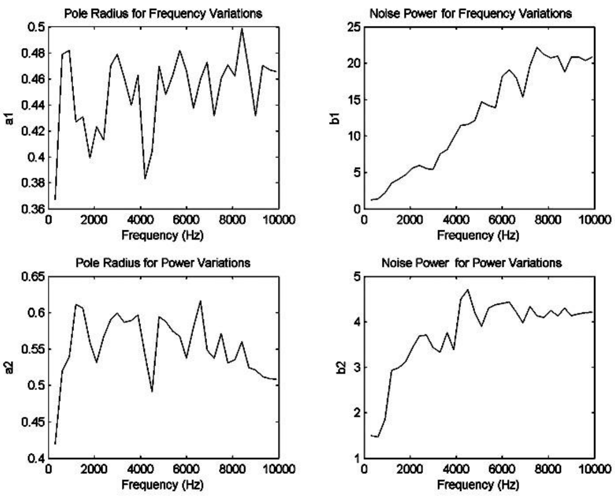

Figure 3: *Estimated parameters for the 4th order model.*

4. ROBUST TRACKING

As mentioned earlier, in practical applications, where parameters of the evolution model are not guaranteed to be accurate enough, the performance of Kalman filter can be poor. The conventional Kalman tracker that we proposed before [6] and that of [4] are not exempt from this limitation. With the same model parameters for different instruments in one class, we ended up with more false tracks where we were dealing with smoother partials, and we got more missing tracks where we had partials with larger variations.

In addition to the inaccuracy of these parameters, a large amount of effort is needed for their estimation. By a close look at the estimated values for our parameters in Figure 3, one can realize that pole radii for both frequency and power vary arbitrarily and are bounded in small intervals. As a solution to these problems we use a regularized least squares Kalman tracker in the class of robust Kalman filters [8]. This filter promises significant enhancements over the conventional tracker and other partial tracking methodologies. The computationally complexity of this tracker can be noticeably reduced through some useful approximations that do not affect its tracking performance. This discussion will fallow after formulation of the filter and in section 5.1.

4.1. Regularized Least Squares Kalman Tracker

This class of robust Kalman filtering is motivated by estimation techniques for solution of regularized least-squares problems. Compared with the standard Kalman filter, which minimizes the regularized residual norm at each iteration, this filter is designed to minimize the worst-possible regularized residual norm over the class of admissible uncertainties at each iteration [8].

Consider a state-space description of the form

$$\begin{aligned} x(k+1) &= \left(A+\delta A\right)x(k)+Bv(k) \\ y(k) &= Cx(k)+w(k) \end{aligned} \tag{7}$$

where $\{x(0), v(k), w(k)\}$ are uncorrelated zero-mean random variables with covariance matrices Π, Q and R respectively. The perturbation of A is modeled as

$$\delta A = D\Delta E \tag{8}$$

for some known matrices $\{D, E\}$ and for an arbitrary Δ, $\|\Delta\| \le 1$. Then the recursive formulation for our robust tracker can be written as

$$\begin{aligned} M(k) &= \widehat{P}(k/k-1)C^T\left[\widehat{R}+C\widehat{P}(k/k-1)C^T\right]^{-1} \\ \hat{x}(k/k) &= \hat{x}(k/k-1)+M(k)\left[y(k)-C\hat{x}(k/k-1)\right] \\ P(k/k) &= \left[I-M(k)C\right]\widehat{P}(k/k-1) \\ \hat{x}(k+1/k) &= \widehat{A}\hat{x}(k/k)+Bu(k) \\ P(k+1/k) &= \widehat{A}P(k/k)\widehat{A}^T+BQB^T \end{aligned} \tag{9}$$

where

$$\begin{aligned} \widehat{R} &= R-\hat{\lambda}^{-1}CDD^TC^T \\ \widehat{P}(k/k-1) &= \left(P^{-1}(k/k-1)+\hat{\lambda}E^TE\right)^{-1} \\ \widehat{A} &= A\left(I-\hat{\lambda}\widehat{P}(k/k-1)E^TE\right) \end{aligned} \tag{10}$$

Here, $\hat{\lambda}$ is a nonnegative scalar parameter that can be determined from optimization

$$\hat{\lambda} = \arg \min_{\lambda \ge \|H^TWH\|} G(\lambda) \tag{11}$$

where the function $G(\lambda)$ is defined as

$$\begin{aligned} G(\lambda) = {} & z^T(\lambda)Q_f z(\lambda)+\lambda\left\|Ez(\lambda)+E\hat{x}(k/k-1)\right\|^2 \\ & +\left(Fz(\lambda)-b\right)^T W(\lambda)\left(Fz(\lambda)-b\right) \end{aligned} \tag{12}$$

and

$$\begin{aligned} W(\lambda) &\triangleq W+WH(\lambda I-H^TWH)^{-1}H^TW \\ Q(\lambda) &\triangleq Q_f-\lambda E^TE\hat{x}(k/k-1) \\ z(\lambda) &\triangleq \left[Q(\lambda)+F^TW(\lambda)F\right]^{-1} \\ & \quad \times\left[F^TW(\lambda)b-\lambda E^TE\hat{x}(k/k-1)\right] \end{aligned} \tag{13}$$

The relation between new parameters in (11)-(13) and those in (9) and (10) is as follows

$$\begin{aligned} b &= y(k+1)-CA\hat{x}(k/k-1) \\ F &= C\left[A \quad B\right] \\ Q_f &= \left(P^{-1}(k/k-1)\oplus Q^{-1}\right) \\ W &= R^{-1} \\ H &= CD \end{aligned} \tag{14}$$

Here, the notation $(a \oplus b)$ denotes a block diagonal matrix with entries *a* and *b*. The minimization of (11) will always yield a unique solution for $\hat{\lambda}$, since $G(\lambda)$ will always have a global minimum in the interval $\left[\left\|H^T WH\right\|, \infty\right)$ [11]. For convenience we can denote the lower bound of λ by λ_l, where

$$\lambda_l \triangleq \left\|H^T WH\right\| \tag{15}$$

Based on observations from our simulations and others [8], the function $G(\lambda)$ usually reaches its minimum at values that are very close to λ_l. This useful observation suggests that instead of lengthy calculations for finding minimum of $G(\lambda)$, we can use a practical approximation for finding $\hat{\lambda}$. This approximation can be of the form

$$\hat{\lambda} = (1+\alpha)\lambda_l \tag{16}$$

This discussion is further elaborated in section 5.1. For now we use (11) to find $\hat{\lambda}$.

4.2. Tracking Procedure

Our robust tracker is initiated with peak data from the first frame, with the initial values

$$\begin{aligned} \hat{x}(1/0) &= \left[f_i(0) \quad p_i(0) \quad 0 \quad 0\right]^T \\ P(1/0) &= \left(\Pi^{-1} + C^T R^{-1} C\right)^{-1} \end{aligned} \tag{17}$$

and

$$\begin{aligned} Q &= \Pi = I \\ R &= \begin{bmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_p^2 \end{bmatrix} \end{aligned} \tag{18}$$

where σ_f^2 and σ_p^2 are the variances of observation noise processes, and take values close to one. This procedure is presented for the CEI model of (5), and for DEI model we have one added zero state in (17).

At the next step we calculate $\hat{\lambda}$ by minimizing $G(\lambda)$ numerically over the interval $\left[\left\|H^T WH\right\|, \infty\right)$ and this process is repeated for every iteration. After fining $\hat{\lambda}$ we use modifications of (10) and Kalman tracker of (9) to estimate noise-free values for power and frequency in the following frame. If the following frame contains a peak that is close enough to the estimated peak, that peak is added to the track and is used to update the tracker. This process is continued through successive frames until there is no peak close enough to the last estimated peak. Here, the track is terminated or considered as "dead" and a new track is initiated in the following frame. The process starts with all peaks in the first frame and also with all peaks from other frames that have not been used in any track.

4.3. Adaptive Acceptance Gate

A peak is close enough to our estimated peak if it falls into the acceptance gate of the track. We use the distance function or *Mahalanobis* distance [12] to define the closeness of peaks to the estimated values as follows

$$d^2(k) = e^T(k)\left[CP(k/k)C^T + \hat{R}\right]^{-1} e(k) \tag{19}$$

$$e(k) = y(k) - C\hat{x}(k/k) \tag{20}$$

Here, $e(k)$ is the error between current observation and the predicted values, and $CP(k/k)C^T + \hat{R}$ is the covariance matrix of this error. A peak falls into the acceptance gate of an estimated peak if the value of its distance function is less that the gate value. If more than one peak is in the acceptance gate, the one with less distance is selected.

Based on our experience, if we set a universal value for our acceptance gate, the tracking result will be poor. The nature of our frequency tracks is suggestive of an adaptive acceptance gate with different values at different frequencies. As mentioned earlier, we are dealing with pseudo-stationary signals. Frequencies of our partials vary with time but these variations are magnified when we move from lower harmonics to the higher harmonics. So, if we consider the same value for our acceptance gate in all frequencies, we have the risk of missing tracks in higher frequencies or loosely accept false partial tracks in lower frequencies. To cope with these variations we set the gate value as a function of frequency, which is

$$g(f) = 10 + 0.01f \tag{21}$$

In fact, we increase the chance of continuing a track where the peaks are sparser and less likely to join a track with lower variations.

4.4. Missing Peaks

Due to imperfections in estimating the spectrum and also because partials with low power can get buried in noise, we might face the problem of missing peaks. This can result in discontinuities in parts of a partial. To overcome this problem, it is proposed in [1] to add "zombie" states to the end of a track where we cannot find any peak within the acceptance gate. In our algorithm we update the track with estimated states in such situation, and continue this process for a maximum of three frames. If during these attempts no peak falls into the acceptance gate, we consider that track as dead and extract the fake updates from the track. If we find a peak during this process, the track is updated with this peak and we keep the fake updates or zombies.

4.5. Backward Tracking

To add to the accuracy of our algorithm we can perform a backward tracking at the end of each track. When a track is terminated, we can initiate a backward tracker with the last updated states and error covariance matrix. This process is identical to the forward tracking but in the reverse direction. This can be helpful because the forward tracker, especially for the case of the conventional Kalman tracker, is loosely initiated with the noisy observations for power and frequency and zero values for other states, while our backward tracker is initiated more accurately. On the other hand, the backward tracker is capable recovering discontinuity in the forward tracking results, since it has the support of a more accurate initiation and a longer history of observation updates. An example of tracking improvements in using the backward tracker for the case of our conventional tracker is presented in [6].

4.6. Crossing and Closely-Spaced Partials

Although power and frequency partials evolve independently from each other, considering a function of both power and frequency for the distance function in (19) is especially rewarding when we are dealing with crossing or closely-spaced partials. This can mostly happen in the polyphonic scenarios where we can have a combination of either constant and linearly decaying power partials or closely-spaced frequency partials.

In partial tracking techniques, where power and frequency partials are tracked separately ([1], [2], [3]), the problem of crossing partials needs considerable attention and requires additional adjustments to the original tracker. However, in our conventional tracker of [6] and our robust algorithm here the contribution of constant and distinct frequency partials in the distance function helps the tracker to distinguish between the corresponding power partials in the crossing area, and it does not need additional adjustments. An example of this case is presented in [6], while our robust tracker also enjoys the same property.

In the same fashion, closely spaced frequency partials can be successfully tracked with the support of sparsely-spaced power partials. This property is more prominent in the robust tracker than the conventional tracker of [6].

5. RESULTS

To use our tracker we first need to put our models in (5) and (6) into an appropriate form considered in (7) and (8). This process is presented for the forth order model in (7) only. Intending to move uncertainties in the input matrix into the transition matrix, we write

$$A + D\Delta E = \begin{bmatrix} 1 & 0 & \tilde{b}_1 & 0 \\ 0 & 1 & 0 & \tilde{b}_2 \\ 0 & 0 & a_1 & 0 \\ 0 & 0 & 0 & a_2 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \hat{b}_1 & 0 \\ 0 & \hat{b}_2 \end{bmatrix} \tag{22}$$

where

$$b_1 = \tilde{b}_1 \hat{b}_1$$

$$b_2 = \tilde{b}_2 \hat{b}_2$$

For isolating bounded uncertainties, which appear in Figure 3, into Δ, we can write

$$\tilde{b}_1\hat{b}_1 = [2, 20], \; \hat{b}_1 = 5$$

$$\tilde{b}_1 = [0.4, 4] = 2.2 + 1.8 \times [-1, 1] = 2.2 + 1.8\delta_1$$

$$\tilde{b}_2\hat{b}_2 = [1.5, 4.5], \; \hat{b}_2 = 3$$

$$\tilde{b}_2 = [0.5, 1.5] = 1 + 0.5 \times [-1, 1] = 1 + 0.5\delta_2$$

$$a_1 = [0.36, 0.5] = 0.43 + 0.07 \times [-1, 1] = 0.43 + 0.07\delta_3$$

$$a_2 = [0.43, 0.61] = 0.52 + 0.09 \times [-1, 1] = 0.52 + 0.09\delta_4$$

which results in

$$A = \begin{bmatrix} 1 & 0 & 2.2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0.43 & 0 \\ 0 & 0 & 0 & 0.52 \end{bmatrix}, E = \begin{bmatrix} 0 & 0 & 1.8 & 0 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0.07 & 0 \\ 0 & 0 & 0 & 0.09 \end{bmatrix} \tag{23}$$

$$D = I_4, \;\; \Delta = diag(\delta_i), i = 1, ..., 4 \tag{24}$$

We examined the accuracy of our algorithm by performing the proposed partial tracking on a wide range of instrumental sounds from different classes of melodic instruments. The tracking results were compared with our conventional tracker [6] as well as the method proposed in [4]. This comparison was done by first defining some accuracy factors. These factors are

$$R_{dt} = \frac{n_{dt}}{n_{et}} \times 100, \qquad R_{ft} = \frac{n_{ft}}{n_{et}} \times 100 \tag{25}$$

where R_{dt} is the detection rate, R_{ft} is the false rate, n_{dt} is the number of detected tracks, n_{ft} is the number of false tracks, and n_{et} is the number of expected tracks. We computed these factors for 32 musical notes (about 450 partials) from all classes of melodic instruments. The comparison results are shown in Table 1.

Method	R_{dt}	R_{ft}
Robust Kalman Tracker	98.3	9.4
Conventional Kalman Tracker	98.2	18.2
Method of [4]	84.7	27.4

Table 1: *Tracking accuracy rates*

As compared with the method of [4], our conventional and robust tracker do a better job in detecting more real tracks and forming less false tracks. This is mainly because of our adaptive acceptance gate, while in [4] a constant acceptance gate (i.e. 10) is considered for all frequency regions.

Furthermore, the superior performance of the robust tracker is evident in its higher detection rate and significantly lower false rate. This is mostly due to the robustness properties of the robust Kalman tracker. This can be further observed in Figure 4, where our robust estimates track the frequency partial more closely than the conventional tracker. Since in the conventional tracker of [6] and [4] estimated and averaged values for parameters of the evolution models are used, the deviation of estimates can be randomly high and divertive. These deviations can increase the risk of forming false partial tracks.

In polyphonic settings, where we can have more than one musical note at a time, harmonics of different notes can get very close to each other. In this situation, estimates with higher deviations can follow the wrong trajectory (see Figure 5). In the conventional tracker this is due to the more weight given to noise power in higher frequencies for coping with magnified frequency variations in higher frequencies (see the right side of Figure 3). On the contrary, the tracking properties of our robust tracker are not influenced by these variations.

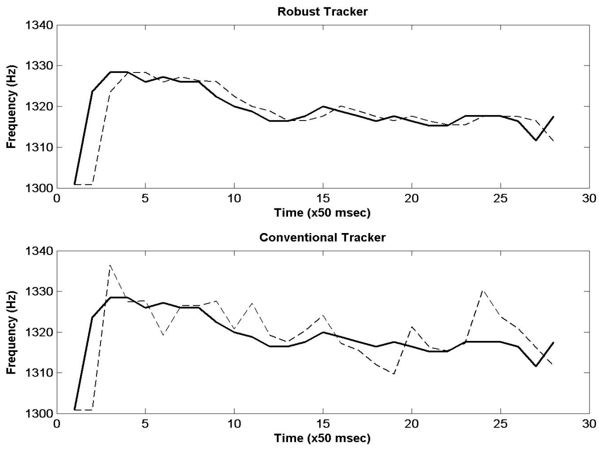

Figure 4: *Performance of the two trackers. Upper: Robust tracker with estimated (dashed) and observed (solid) tracks. Lower: conventional tracker with estimated (dashed) and observed (solid) tracks.*

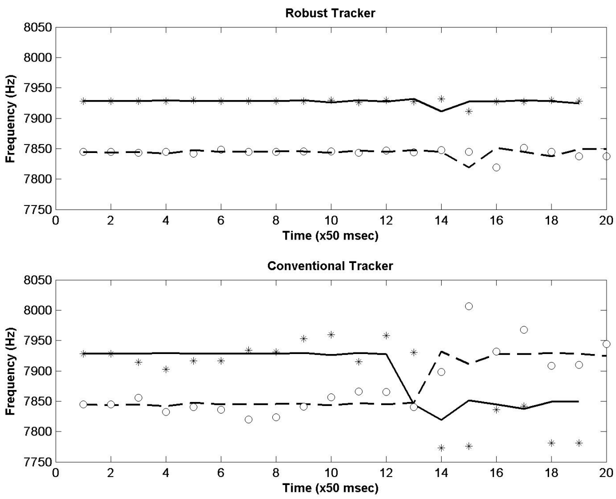

Figure 5: *Tracking in the presence of closely spaced partials: Robust tracker (upper) and conventional tracker (lower); with observed track for higher harmonic (solid), its estimates (asterisk), observed track for lower harmonic (dashed) and its estimates (circle).*

5.1. Useful approximation

Considering (14), (15), (18), and (24) we have

$$\lambda_l = \max(\sigma_f^{-2}, \sigma_p^{-2}) \tag{26}$$

If $\sigma_f^2 = \sigma_p^2 = 0.97$ then we get λ_l =1.031. In concurrence with observations of [11], through all of our experiments the calculated values of $\hat{\lambda}$ in (11) were very close to the lower bound λ_l. This observation suggests using an approximation for this parameter as indicated in (16), which can reduce the computational expense of our algorithm significantly. In fact, using $\hat{\lambda} = 1.1$ did not have any considerable effect on the tracking results throughout our experiments, and our robust tracker preserved all its prominent properties.

6. CONCLUSION

We presented a novel partial tracking method based on a robust Kalman filter. The tracker displays improved capabilities in tracking partials thanks to its robustness properties. The computational complexity of the recursive algorithm can be noticeably reduced by a practical approximation. As the continuation of this work we can further investigate the possibility of using a universal evolution model for all classes of melodic instruments.

7. REFERENCES

[1] X. Serra, "Musical Sound Modelling with Sinusoids plus Noise," in *Musical Signal Processing*, Lisse [Netherlands]; Exton, PA: Swets & Zeitlinger, 1997, pp. 91-122.

[2] R. McAulay and T. Quatieri, "Speech analysis/Synthesis based on a sinusoidal representation," *Acoustics, Speech, and Signal Processing [see also IEEE Trans. Signal Processing], IEEE Trans.*, vol. 34, pp. 744-754, 1986.

[3] M. Lagrange, S. Marchand, and J. B. Rault, "Using Linear Prediction to Enhance the Tracking of Partials," presented at Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04). IEEE International Conference on, 2004.

[4] A. Sterian, "Model-Based Segmentation of Time-Frequency Images for Musical Transcription," in *Electrical Engineering: Systems*. Ann Arbor: University of Michigan, 1999.

[5] W. Roguet, N. Martin, and A. Chehikian, "Tracking of frequency in a time-frequency representation," presented at Time-Frequency and Time-Scale Analysis, Proceedings of the IEEE-SP International Symposium on, 1996.

[6] H. Satar-Boroujeni and B. Shafai, "Kalman Filtering Application in Automatic Music Transcription," accepted for the 2005 IEEE Conference on Control Applications, Toronto, Canada, 2005.

[7] Y. Theodor, U. Shaked, and C. E. de Souza, "A Game Theory Approach to Robust Discrete-time H-infinity Estimation," *IEEE Trans. Signal Processing*, pp. 1486–1495, 1994.

[8] A. H. Sayed, "A framework for state-space estimation with uncertain models," *Automatic Control, IEEE Transactions on*, vol. 46, pp. 998-1013, 2001.

[9] H. Satar-Boroujeni and B. Shafai, "Peak Tracking and Partial Formation of Music Signals," accepted for the 13th European Signal Processing Conference, Antalya, Turkey, 2005.

[10] H. Satar-Boroujeni and B. Shafai, "State-Space Modeling and Analysis for Partial Tracking of Music Signals," presented at The 24th IASTED International Conference on Modelling, Identification, and Control, Innsbruck, Austria, 2005.

[11] A. H. Sayed and H. Chen, "A uniqueness result concerning a robust regularized least-squares solution," *Systems & Control Letters*, vol. 46, pp. 361-369, 2002.

[12] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. New York: Wiley, 2001.