Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Robust Kalman Filtering with Application to Tracking of Partials in Music Signals

Hamid Satar-Boroujeni, *Member, IEEE,* and Bahram Shafai, *Senior Member, IEEE*

***Abstract*—In this paper we propose a novel method for tracking of partials in music signals based on a robust Kalman filter. This 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. This method promises improved tracking capabilities, compared with the conventional Kalman filter, which was proposed before. The model parameters that have been estimated for different frequencies are now considered as bounded uncertainties. Unlike the conventional Kalman tracker, the performance of this tracker is not influenced by the magnified track variations in higher frequencies.**

I. 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 (e.g. music analysis/synthesis [1], automatic music transcription [2], audio restoration [3], etc.) 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 small frames of the temporal data that can be assumed to be 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 [4] and then adopted for the case of music signals [1], where it was based on a heuristic approach. In a more recent approach [5] and as an extension to [4], 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 behaviour of peaks' amplitude is not taken into account while performing the tracking. Another approach [2], which was inspired by a similar technique in radar tracking and also a frequency tracker for avalanche signals [6], takes the advantage of Kalman filter by constructing a state-space model for the behaviour 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 behaviour of a peak's frequency and power.

We proposed a partial tracking technique before [7], 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 [8].

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).

This paper will proceed with an introduction to automatic music transcription as a main application area for our partial tracker. In section III 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 [9], will be discussed in section IV. In section V we will include some results and compare the performance of this tracker with the conventional Kalman tracker.

II. Background on Music Signal Analysis

A. Automatic Music Transcription

Music transcription is the process of un-making or documenting music. Un-making in the sense that the process of reading from score and playing music is reversed [2], and documenting in the context of substantiating musical sounds, whether it has been played from score, memory, or just improvised.

A music transcription system, in its perfection, should be able to detect all attributes as written in the score, such as loudness and tempo, as well as performance gestures intended by the performer. However, at the fundamental level it is the problem of recognizing which note is played and when. This process has been done by human listener with trained ears, but developing a music transcription system, which replaces the human listener with a computer,

The authors are with Electrical and Computer Engineering department, Northeastern University, Boston, MA 02115, USA (phone: 617-373-2984; fax: 617-373-4189; e-mail: {hsattar&shafai}@ ece.neu.edu).

even at the basic level, requires sophisticated signal processing techniques. For polyphonic music, where more than one musical note is present at a time, keeping track of individual notes and producing the simple score is a more challenging matter.

Each musical note contains a fundamental frequency and integer multiples of this frequency, which are called harmonics of the note. To identify a note we need to know its fundamental frequency or pitch. For transcribing a piece of music, the identity of each note and its time duration are required. Since in a real scenario we can have more than one note being played at a time, we need to distinguish between these notes. This can be done by identifying all the partials and their initiation and termination times first. The fundamental and all the harmonics related to each note and their time length are then extracted, which can be directly interpreted into musical score.

A general music transcription system takes the waveform of recorded music and finds the behavior of frequencies within small time frames using spectral estimation tools and assuming that the signal is a combination of sinusoids and noise. In fact, we are dealing with pseudo-stationary signals for which amplitude and frequencies vary slowly with time, and we choose small time frames to preserve the stationarity which is required for the estimation of spectrum. This process results in a representation with power concentrated at specific frequencies. These frequencies, which are the local maxima within the spectral representation, are indications of partials of existing musical notes in that time frame. Identified peaks from adjacent frame which belong to the same partial must be connected to each other using data association techniques. One possible approach is to use Kalman filtering technique for tracking partials through neighboring time frames based on a pre-estimated state-space model for evolution of frequency and power in time. This idea is pretty much equivalent to the use of Kalman filtering in the area of radar tracking [10].

In the process of music signal analysis, detection of peaks plays an important role. We need to collect all possible peaks pertaining to existing partials and reject 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. Having this in mind we proposed a novel and improved technique for detection of peaks, which was introduced before **Error! Reference source not found.**.

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

III. State-Space Modeling

A. 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

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

Here, $s(t)$ reflects the pure musical part of the signal and $e(t)$ 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 $\varphi_n(t)$ represents timbral variations and performance effects. Since we do not consider such effects in our music signals, $\varphi_n(t)$ will be considered as a noise process.

B. 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, in fact, 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, the 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 [12]. 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 will be in the from of

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

We can estimate the parameters of each model, e.g. a_1, b_1, a_2, and b_2, by performing a statistical analysis on a large number of musical sounds with known identities in a forward-problem setting. The details of this procedure are presented in [12].

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 both CEI and DEI classes are shown in fig. 1. In section V, where we transform our model to a form appropriate for robust Kalman tracker, these parameters are treated as bounded uncertainties.

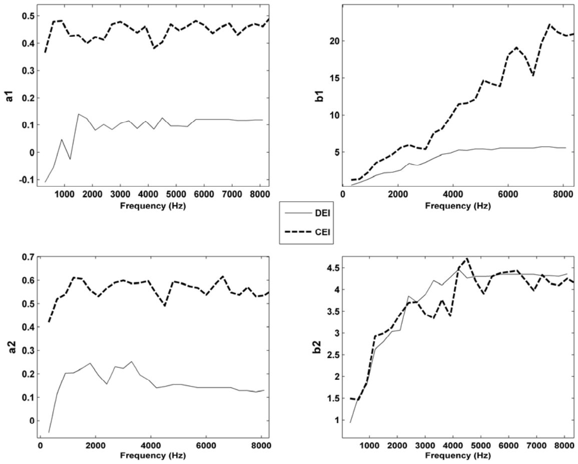

Figure 1: Estimated parameters for both CEI (dashed) and DEI (solid) models.

IV. Robust Kalman Tracker

A. Motivation and Possibilities

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. Our conventional Kalman tracker is 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 fig. 1, one can realize that pole radii for both frequency and power vary arbitrarily and are bounded in small intervals. This motivated us to look for robust trackers where these frequency-varying parameters can be treated as bounded uncertainties. Some possible solutions in the field of robust filtering are as follows:

1) Discrete-Time Quadratic Guaranteed Cost Filtering

This class of filtering is applicable to systems that are quadratically stable [13]. However, our transition matrices in both systems of (5) and (6) have two poles on the unit circle. This means that our systems are not Schur stable and do not satisfy the conditions for quadratic stability.

2) Discrete-Time Set-Valued State Estimation [13], Robust H-infinity Filtering [8], etc.

For these classes of robust filtering we face the same limitation as the first option: our system must by Schur stable. Therefore, we can not employ these methods either.

3) Regularized Least Squares Approach [9].

This class of robust Kalman filtering, which is motivated by estimation techniques for solution of regularized least-squares problems, does not need the stability condition. In the next subsection we formulate a robust Kalman tracker based on this approach.

B. Regularized Least Squares Kalman Tracker

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 [9].

Consider a state-space description of the form

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

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

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

where

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

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 $[\|H^TWH\|, \infty)$ [14]. For convenience we can denote the lower bound of λ by λ_l, where

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

Based on observations from our simulations and others [9], 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 V. For now we use (11) to find $\hat{\lambda}$.

C. Tracking Procedure

Our robust tracker is initiated with peak data from the first frame, with the initial values (in the forth order model)

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

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. At the next step we calculate $\hat{\lambda}$ by minimizing $G(\lambda)$ over the interval $[\|H^TWH\|, \infty)$ and this process is repeated for every iteration. After finding $\hat{\lambda}$ 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 next 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 previous track.

D. 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* to define the closeness of peaks to the estimated values as follows [10]:

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

Here, $e(k) = y(k) - C\hat{x}(k/k)$ 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

$$g(f) = 10 + 0.01f \tag{20}$$

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.

E. 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.

V. 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

$$A + D\Delta E = \begin{bmatrix} 1 & 0 & \tilde{b}_1 & 0 \\ 0 & 1 & 0 & \tilde{b}_2 \\ 0 & 0 & a_1 & 0 \\ 0 & 0 & 0 & a_2 \end{bmatrix},\ B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \hat{b}_1 & 0 \\ 0 & \hat{b}_2 \end{bmatrix} \tag{21}$$

where

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

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

For isolating bounded uncertainties, which appear in fig. 1, 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

$$A = \begin{bmatrix} 1 & 0 & 2.2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0.43 & 0 \\ 0 & 0 & 0 & 0.52 \end{bmatrix},\ E = \begin{bmatrix} 0 & 0 & 1.8 & 0 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0.07 & 0 \\ 0 & 0 & 0 & 0.09 \end{bmatrix} \tag{22}$$

$$D = I_4,\ \ \Delta = diag(\delta_i), i = 1, ..., 4 \tag{23}$$

Considering (14), (15), (18), and (23) we have:

$$\lambda_l = \max(\sigma_f^{-2}, \sigma_p^{-2}) \tag{24}$$

If $\sigma^2_f = \sigma^2_p$=0.97 then we get λ_l=1.031. In concurrence with the observations of [9], through all of our simulations 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. The study of this case is left for future work.

The robust tracker is proposed as an improvement to our earlier tracker, which was based on conventional Kalman filter. To compare the performance of these two trackers we use the accuracy factors introduced in [7], which 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. Table 1 contains accuracy rates for these two trackers.

The superior performance of the robust tracker is evident in its higher detection rate and significantly lower false rate, compared with the conventional tracker. This is mostly due to robustness of the new method and its improved tracking

Method	R_{dt}	R_{ft}
Conventional Kalman Tracker	98.2	18.2
Robust Kalman Tracker	98.3	9.4

Table1: Tracking accuracy rates

capabilities. This improvement can be further observed in fig. 3, where our robust estimates track the frequency partial more closely than the conventional estimates. Since we used estimated and averaged values for parameters of the evolution models in our conventional tracker, the deviation of estimates can be randomly high and divertive. This deviations increase the risk of forming false partial tracks.

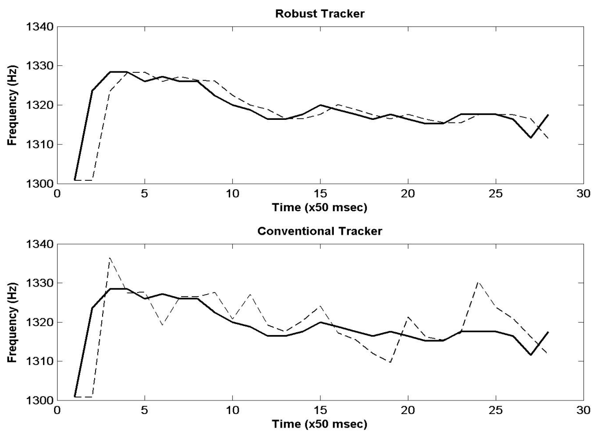

Figure 3: Performance of the two trackers. Upper: Robust tracker with estimated track (dashed) and observed track (solid). Lower: conventional tracker with estimated track (dashed) and observed track (solid)

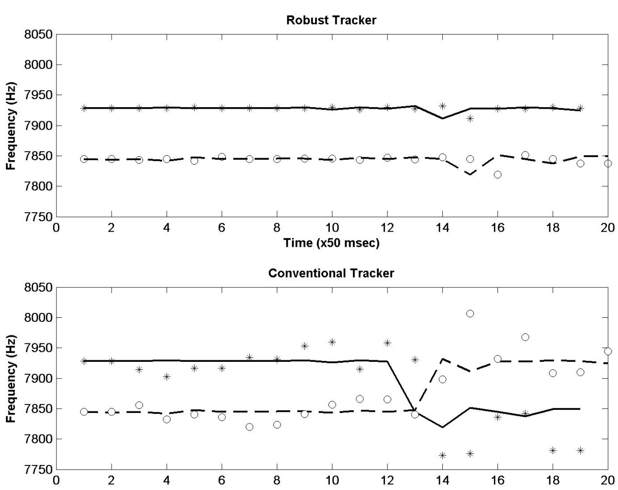

Figure 4: Tracking in the presence of closely spaced partials: Robust tracker (upper) and conventional tracker (lower); with observed track for higher harmonic (solid), its estimates (asterisk), observed track for lower harmonic (dashed) and its estimates (circle).

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 fig. 4). In the conventional tracker this is due to the more weight given to noise power in higher frequencies (see the right side of fig. 1) for coping with magnified frequency variations in higher frequencies. On the contrary, the tracking properties of our robust tracker are not influenced by these variations.

VI. Conclusion

We presented a new partial tracking method based on a robust Kalman filter. This tracker promises an improved performance over our conventional Kalman tracker while preserving its good properties and superiority over existing methodologies. As the continuation of this work we can further investigate the possibility of using a universal evolution model for all classes of melodic instruments.

References

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

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

[3] S. J. Godsill and P. J. W. Rayner, *Digital audio restoration : a statistical model based approach*. London; New York: Springer, 1998.

[4] 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.

[5] 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.

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

[7] H. Satar-Boroujeni and B. Shafai, “Kalman Filtering Application in Automatic Music Transcription,” presented at IEEE Conference on Control Applications, Toronto, Canada 2005.

[8] 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.

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

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

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

[12] 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.

[13] I. R. Petersen and A. V. Savkin, *Robust Kalman filtering for signals and systems with large uncertainties*. Boston: Birkhäuser, 1999.

[14] 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.