Improved Position Tracking of a 3-D Gesture-Based Musical Controller Using a Kalman Filter

Manjinder Singh Benning[1, 3] Michael McGuire[2] Peter Driessen[1]
mbenning@ece.uvic.ca mmcguire@ece.uvic.ca peter@ece.uvic.ca

[1]Music Intelligence and Sound Interdisciplinary Centre (MISTIC), University of Victoria, BC, Canada

[2]Digital Signal Processing Group, University of Victoria, BC, Canada

[3]KarmetiK Technology (A Division of KarmetiK LLC), Reno, NV, USA

ABSTRACT

This paper describes the design and experimentation of a Kalman Filter used to improve position tracking of a 3-D gesture-based musical controller known as the Radiodrum. The Singer dynamic model for target tracking is used to describe the evolution of a Radiodrum's stick position in time. The autocorrelation time constant of a gesture's acceleration and the variance of the gesture acceleration are used to tune the model to various performance modes. Multiple Kalman Filters tuned to each gesture type are run in parallel and an Interacting Multiple Model (IMM) is implemented to decide on the best combination of filter outputs to track the current gesture. Our goal is to accurately track Radiodrum gestures through noisy measurement signals.

Keywords

Kalman Filtering, Radiodrum, Gesture Tracking, Interacting Multiple Model

1. INTRODUCTION

Intention is a key aspect of traditional music performance. The ability for an artist to reliably reproduce sound, pitch, rhythms, and emotion is paramount to the design of any instrument. With the introduction of acoustically quite electronic musical controllers, intention is determined by the accuracy of the sensor technology and engineering of the gesture capturing system.

Most controllers are susceptible to a fair amount of unpredictable electromagnetic noise either from with in the sensing system itself or from the surrounding environment. An accurate track of the gesture data coming from a multi-dimensional performance based sensing system is useful for many applications. Accurate estimation of the true gesture provides a performer with greater control of the virtual space. The Tongue 'n' Groove ultrasound based controller discusses this need for increased accuracy of data [8]. Furthermore, recognition in gesture based conducting systems such as [9, 11, 12], would benefit from an improved gesture track.

The motivation for this paper is to process the gesture data measured by our sensing system and reduce, if not completely remove all influence of system and environmental noise on the gesture signals. With an understanding of our system's sensor noise and models of the various gesture motion types expected, a filter can track our multi-dimensional gesture signal through the noisy raw signal. The Kalman Filter is an excellent candidate for such a problem.

A similar problem was addressed in data correction for a system that tracked 'air percussion' gestures [7]. In this work the author uses LPC prediction, with a window of 30 samples, to predict the next sample and smoothes the measured data by averaging with the predicted data point. Our work differs in that we use a model-based approach to predict the next sample and the weighting between the measured and predicted data point is conditional on all previous measurements.

This paper describes the design of Kalman Filters to improve the tracking of the Radiodrum system. Section 2 describes the Radiodrum system, section 3 discusses the dynamic model used to describe the motion of a drumstick and section 4 describes a measurement model of the Radiodrum system. Section 5 describes the 3 Kalman Filters used to track each gesture type while section 6 discusses the implementation and results of an Interactive Multiple Model used to combine our Kalman Filter outputs. Section 7 contains conclusions and future work.

2. THE RADIODRUM

Also often referred to as the Radio Baton, The Radiodrum is a 3 dimensional musical controller that tracks the x, y, and z position, z velocity and detects surface whacks of one or two drum sticks over its surface. Originally designed and built at Bell Laboratories in the 1980's to be used as a 3 dimensional mouse, the Radiodrum has now evolved to become a pioneering instrument in computer music performance [6, 10].

The current Radiodrum system developed by Ben Neville [1] uses an audio interface to generate the emitted carrier signals and to acquire the antenna signals coming off the Radiodrum surface. In Max/MSP, the four antenna signals are demodulated and translated into x, y, and z positions. In our work we will limit ourselves to a single stick system. The results of this research can easily be extended to work with a second stick. Figure 1 shows a 2 second record of the x position of the Radiodrum stick moving slowly over the surface. Here, the undesirable effects of the noise are obvious. At the noisiest part of the signal, the exact stick position in the x direction is ambiguous in the range of 20-30cm.

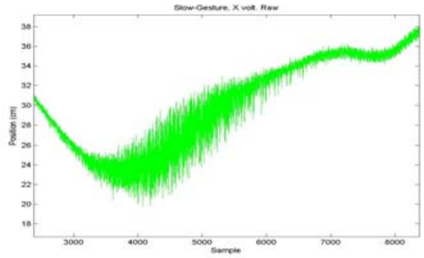

Figure 1: Raw x position of Radiodrum signal

3. DYNAMICAL SYSTEM FOR RADIODRUM GESTURES

Motion of a Radiodrum stick can be described by a linear dynamic model consisting of a state vector X(k) and a state transition matrix Φ(k). Random disturbances to the state due to the randomness of human motion can be modeled by a white Gaussian noise process, W(k), described by its covariance matrix, Q(k). Equation 1 describes the state vector containing the position, velocity, and acceleration of a Radiodrum stick in all three dimensions.

$$X(k) = [x, vx, ax, y, vy, ay, z, vz, az]$$

Equation 1: State Vector

Equation 2 describes how the system's state evolves from one time instant to the next.

$$X(k+1) = \Phi(k)X(k) + W(k)$$

Equation 2: Linear Dynamic Model

Φ(k) and Q(k) are obtained using the Singer model for maneuvering targets described in section 3.1.

3.1 The Singer Dynamic Model

The dynamic model for the Radiodrum system was inspired by extensive work already published in the field of missile tracking for military applications. Tracking of a Radiodrum stick is analogous to tracking a missile through the earth's atmosphere using radar. The Singer model for maneuvering targets provides a intuitive parameterized way to specify the state transition matrix, Φ(k), and the dynamic disturbance covariance matrix, Q(k) [2].

The Singer model takes into account a dynamic acceleration disturbance noise process that is not necessarily white. Physically, this means that a force may be applied to the stick over a certain window of time to maneuver it. The correlation coefficient τ describes to what degree the noise is correlated and σ^2 is the variance of the acceleration over a gesture. Equation 3 shows the autocorrelation of the acceleration. T is the sampling period. In our case 1/3000 sec.

$$E[a(t)a(t+T)] = \sigma^2 e^{(-\frac{|T|}{\tau})}$$

Equation 3: Autocorrelation of acceleration

The Φ(k) matrix, shown in Equation 4, describes how the state of a single dimension of the Radiodrum is updated over consecutive time instants.

$$\begin{bmatrix} 1 & T & \tau^2[-1+\frac{T}{\tau}+e^{(-\frac{T}{\tau})}] \\ 0 & 1 & \tau[1-e^{(-\frac{T}{\tau})}] \\ 0 & 0 & e^{(-\frac{T}{\tau})} \end{bmatrix}$$

Equation 4: Transition matrix, Φ(k)

Since the random disturbance matrix W(k) must be white the random acceleration component is processed through a whitening filter giving rise to a Q(k) matrix of the form in Equation 5.

$$Q(k) = E[W(k)W^T(k)] = \frac{2\sigma^2}{\tau}\begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix}$$

Equation 5: Covariance of disturbance matrix

The q values are functions of τ, the correlation time and T, the sampling period. For the exact derivation of Q(k) see [3].

4. RADIODRUM MEASUREMENT MODEL

A measurement model describes the relationship between the known measurements and the unknown parameters of a system.

$$Z(k) = H(k)X(k) + V(k)$$

Equation 6: Linear Measurement Model

Where Z(k) denotes the measurement vector, H(k) denotes the observation matrix, X(k) denotes the unknown parameters and V(k) the measurement noise. The matrix R(k) is defined as the covariance of V(k).

In our case, the measurements correspond directly with x, y, z in the state vector therefore H(k) is simply a (3,9) matrix with zeros except in the entries (1,1), (2,4), and (3,7). The covariance of the x, y, and z position noise was measured throughout the entire playable range of the Radiodrum's 3-dimensional surface. Since the noise is position dependant, more specifically, increasing with height of the stick, an averaged sum of all covariance calculations over the surface was used.

5. KALMAN FILTERING OF THE RADIODRUM GESTURES

The Kalman Filter is an optimal recursive linear estimator. With knowledge of the system and measurement devices, all measurements are processed to estimate the desired unknown parameters. The Kalman Filter processes the measurements in a linear fashion minimizing the error between the estimated parameters and the actual parameters [4]. The Kalman Filter outputs a filtered estimate of the state of the system. This estimate is a weighted combination of the measurements and the predicted state provided by the dynamic model. For a complete derivation of the Kalman Filter, see [4].

This initial work on Kalman Filtering of Radiodrum gestures was implemented and tested offline in Matlab.

5.1 Definition of Gesture Types

A single Kalman Filter, using a specific tuned dynamic model for each dimension cannot successfully track the entire range of gestures that a performer can offer. To cover the breadth of acceleration variances and time correlations, three different dynamic models were used. These three models correspond to three modes of gesture: Slow Move, Fast Move, and Whack. The Slow Move and Fast Move gestures are defined as slow and fast varying stick movements over the range of the surface respectively. The Whack gesture corresponds to the striking of the stick upon the surface of the drum. In whack mode the stick experiences the highest accelerations and decelerations, therefore smallest acceleration correlation (impulsive) with greatest variance.

5.2 Tuning of model parameters

A variety of gestures, representing each mode were tracked using the three Kalman Filters. Table 1 shows the x, y, and z values of the model parameters tuned for each gesture mode.

Gesture Mode	**τ (x, y, z) (s)** **T=1/3000 sec**	**σ^2 (x, y, z) (m/s^2)**
Slow Move	100*T, 100*T,100*T	0.1, 0.1, 100
Fast Move	65*T, 65*T, 65*T	800, 800, 80000
Whack	100*T, 100*T, 2*T	70, 70, 1e7

Table 1: Singer Model Parameters for Gesture Modes

As expected the decrease in **τ** from Slow Move to Fast to Whack indicates the movement getting more impulsive and intuitively, the variance of the acceleration increases as the gestures get more erratic. In all cases the z variance is greater than the x and y. This is because performers tend to move more impulsively in the z direction. For the case of the Whack, an extreme variance models the z acceleration while lower variances model the x and y accelerations. Whacking requires very little movement in the x and y directions.

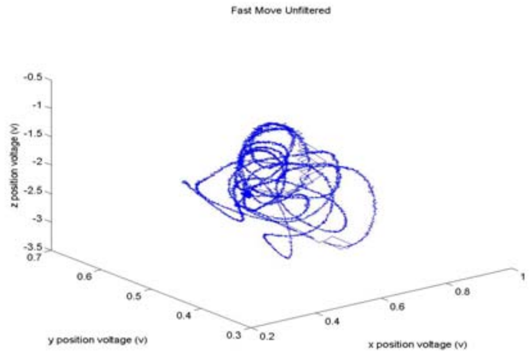

Figure 2: Raw unfiltered Fast Move gesture

Figure 2 and Figure 3 show a 4 second record of the raw and filtered Radiodrum position track of a fast gesture in 3 dimensions respectively. A Kalman Filter tuned to the Fast Move mode was used. This gesture was accurately tracked through noisy measurements with improved accuracy. Next we describe how the 3 models are combined to obtain a single estimate of the current gesture track.

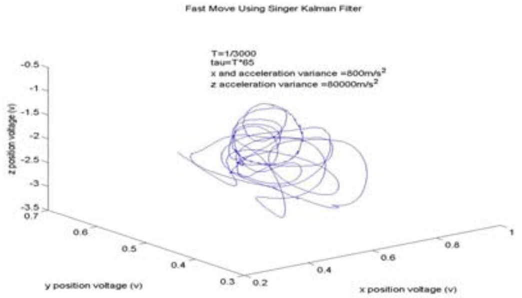

Figure 3: Kalman filtered Fast Move gesture

6. AN INTERACTING MULTIPLE MODEL

The Interacting Multiple Model (IMM) algorithm provides a single output of the estimated state of a dynamic system from a combination of subfilter outputs, each of which are tuned to a specific dynamic model. The weighting of each model's state estimate in the final combined state estimate is proportional to each model's likelihood function at the current iteration of the algorithm. The likelihood of each model, j is calculated from the normal distribution shown in Equation 7 [5].

$$\Lambda_j = \mathrm{N}[z(k); \hat{z}^{j}(k), S^j]$$

Equation 7: Model Likelihood

Where z(k) are the current measurements, $\hat{z}^{j}(k)$ are the current model's predicted measurements, and S^j is the covariance of the model's innovations sequence, the difference between the predicted measurements and actual measurements.

Inputs to each model's filter at the start of an iteration, k, is a weighted combination of the outputs of each model's state estimate, X_j(k-1), and covariance, P_j (k-1) from the previous iteration, k-1. Weighting of each model j's state estimate and covariance is governed by model j's likelihood, Λ_j and model j's switching probability to the current model. Since a Markovian process governs model transition, a model transition probability matrix can be defined.

6.1 IMM Results for the Radiodrum

We use 3 models to describe the 3 modes of performance, differing by their dynamic noise covariance matrix, Q.

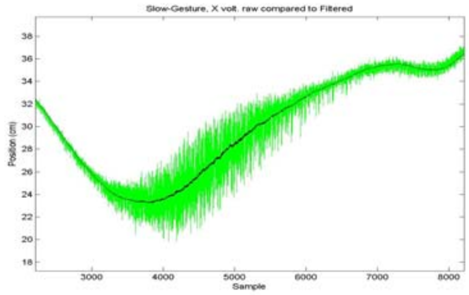

Figure 4: IMM Filtered Radiodrum signal of slow gesture

Figure 4 shows the same record of data as Figure 1. The green line represents the raw unfiltered data obtained from the Radiodrum. The black line running through the green plot shows the filtered output of the 3 model IMM. The filtering has reduced the maximum x position uncertainty of this gesture from 10 cm's to a few millimeters.

Figure 5 plots the raw and corresponding black IMM filtered z track for a slow to fast gesture transition. Although a reasonable position noise reduction from accuracy of 6cm to 1cm is observed, the IMM has trouble estimating the best track for the slow gesture up to sample 3400. The Fast Move and Whack model's outputs are being favoured over the Slow Move model. An IMM favouring the slow model up to sample 3400 would give a position accuracy of a few millimeters. Since the Fast Move and Whack models for the z axis have such large variances, the Slow Move model's distribution gets 'swallowed' and never becomes more likely over the other models.

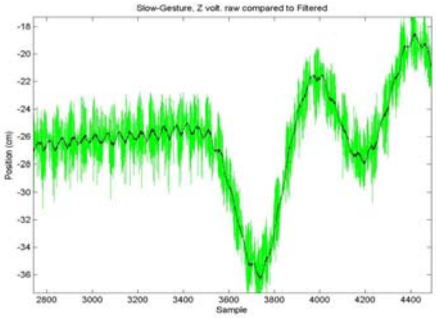

Figure 5: IMM Filtered Radiodrum signal, slow to fast gesture

Any tuning of the current IMM will not give a better track for the z position. A new approach is needed to get a tighter track of z position. Figure 6 exhibits another problem.

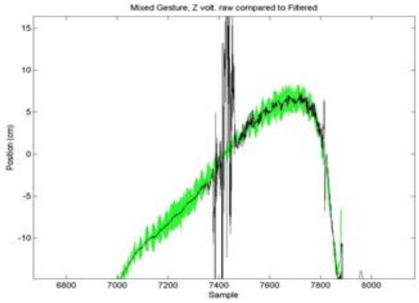

Figure 6: Radiodrum signal noise burst

Here a burst of noise, independent of the gesture, enters the system. The IMM switches from the Slow to the Whack model and closely tracks the noise as if it were gesture. A way is needed to distinguish between noise bursts and gesture.

7. CONCLUSIONS AND FUTURE WORK

Using the Singer model for maneuvering targets, a set of 3 dynamic models, describing 3 modes of Radiodrum gesture were designed. Along with a single measurement model, the states of the 3 models are estimated using 3 unique linear Kalman Filters. Using an Interacting Multiple Model, the 3 Kalman Filter outputs are combined to give a single estimate of the state, and to provide input to the Kalman Filters at the next iteration. The IMM performs well when filtering x and y position at various speeds. However, due to the large variances of the z components for the Fast Move and Whack gestures, the Slow Move model cannot achieve adequate separation in the z direction to be effective. Furthermore, the IMM cannot distinguish between noise bursts and impulsive gesture.

Future work includes filtering each coordinate separately and providing input to the dynamic model in the form of acceleration impulse and stick height. This may provide adequate separation of the distributions of our 3 models and alleviate the problem of falsely tracking noise bursts as whacks. Finally, a Max/MSP based Kalman Filter external will be developed and incorporated into the existing Radiodrum software [1], for real-time performance evaluation.

8. ACKNOWLEDGMENTS

Our thanks to George Tzanetakis, Andrew Schloss, and Ajay Kapur of UVic's MISTIC research group.

9. REFERENCES

[1] Neville, B. *Gesture analysis through a computer's audio interface: The Audio-Input Drum.* MaSc. Thesis, University of Victoria, Victoria, BC, 2006.

[2] E. Brookner, *Tracking and Kalman Filtering Made Easy.* John Wiley and Sons, Inc, New York, USA, 1998.

[3] Singer, R. A. Estimating Optimal Tracking Filter Performance of Manned Maneuvering Targets. *IEEE Transactions On Aerospace and Electronic Systems, Vol AES-6C4*, 1970, pp. 473-484.

[4] J. M. Mendel, *Lessons In Estimation Theory For Signal Processing, Communications, and Control.* Prentice Hall PTR, New Jersey, USA, 1995.

[5] Y. Bar-Shalom, X. Rong Li, T. Kirubarajan. *Estimation with Applications to Tracking and Navigation.* John Wiley and Sons Inc, New York, USA, 2001.

[6] Boie R., Mathews M., and Schloss A. "The Radio Drum as a Synthesizer Controller", *In the Proceedings of the International Computer Music Conference (ICMC)*, San Francisco, USA, 1989.

[7] Goudard V., Havel C., Marchand S., Desainte-Catherine M. "Data Anticipation For Gesture Recognition In The Air Percussion", *(ICMC)*, Barcelona, Spain, 2005.

[8] Vogt F., McCaig G., Mir A, A., Fels S. "Tongue 'n' Groove: An Ultrasound Based Controller", *In the Proceedings of New Instruments for Musical Expression (NIME)*, Dublin, Ireland, 2002.

[9] Bresin R., Hansen K. F., Dahl., S. "The Radio Baton As Configurable Musical Instrument and Controller", *In the Proceedings of the Stockholm Music Acoustic Conference, (SMAC)*, Stockholm, Swedan, 2003.

[10] Boulanger R., Mathews M. "The 1997 Mathews Radio-Baton and Improvisational Modes", *(ICMC)*, Thessaloniki, Greece, 1997.

[11] A. Camurri, P. Coletta, M. Ricchetti, R. Trocca, K Suzuki, and G. Volpe, "Toward a Framework for Interactive Systems to Conduct Digital Audio and Video Streams," *Computer Music Journal,* 30:1, pp 21-36, MIT Press, Spring 2006.

[12] Ilmonen T., Takala M. "Conductor Following With Artificial Neural Networks", *(ICMC)*, San Francisco, USA, 1999