MODELING AND CONTINUOUS SONIFICATION OF AFFORDANCES FOR GESTURE-BASED INTERFACES

Yon Visell and Jeremy Cooperstock

Centre for Intelligent Machines and CIRMMT
McGill University, Montreal, Canada
yon@cim.mcgill.ca jer@cim.mcgill.ca

ABSTRACT

Sonification can play a significant role in facilitating continuous, gesture-based input in closed loop human computer interaction, where it offers the potential to improve the experience of users, making systems easier to use by rendering their inferences more transparent. The interactive system described here provides a number of gestural affordances which may not be apparent to the user through a visual display or other cues, and provides novel means for navigating them with sound or vibrotactile feedback.The approach combines machine learning techniques for understanding a user's gestures, with a method for the auditory display of salient features of the underlying inference process in real time. It uses a particle filter to track multiple hypotheses about a user's input as the latter is unfolding, together with Dynamic Movement Primitives, introduced in work by Schaal et al [1][2], which model a user's gesture as evidence of a nonlinear dynamical system that has given rise to them. The sonification is based upon a presentation of features derived from estimates of the time varying probability that the user's gesture conforms to state trajectories through the ensemble of dynamical systems. We propose mapping constraints for the sonification of time-dependent sampled probability densities. The system is being initially assessed with trial tasks such as a figure reproduction using a multi degree-of-freedom wireless pointing input device, and a handwriting interface.

1. INTRODUCTION

The research discussed here intends to improve the design of gesture-based interactions with the widening range of computational artifacts supporting continuous input. These comprise new computer input devices, video games, musical instruments, and an increasingly ubiquitous class of everyday artifacts that possess hidden computing, sensing, and actuating capabilities. Already familiar examples of the last include running shoes and wireless video game controllers. In the context of human-computer interaction, user experience may be improved if continuous, gesture-based input is situated in complete sensory-motor feedback loops. Non-visual display modalities have some advantages for such closed loop interactions, as they can more easily be situated at the locus of interaction of an arbitrary input device. Other benefits include their ability to provide an effective representation of temporal features and to operate in situations in which the visual modality is overtaxed, both of which have been discussed extensively in previous literature [3][4].

The system described below tracks the state of a user's gestures, as relayed by the sensors of a continuous input device, and continuously estimates a correspondence with a number of gestural affordances. The latter are encoded as nonlinear dynamical systems models that are learned from a gesture dataset, as explained in Section 2.1. A particle filter algorithm is used to track hypotheses as to the correspondence between the user's input and the models. The information provided by this inference mechanism is used to generate a sonification to allow users to guide their gestures relative to those afforded by the system.

1.1. Related work

A particle filter based recognizer for generating sonic feedback to display a user's input in a mouse gestures task, and in related applications such as a helicopter control task, has been described by Williamson and Murray-Smith [5][6]. Their recognizer appears to be built on static template models of the gestures afforded by the system. The system presented here is similar, but uses the dynamical systems models of Schaal et al [1][2], which encode the dynamics of the motion more richly than is possible with templates. The dynamical systems models have an advantage in their ability to representing dynamically driven variations in the detailed path taken while preserving other features of the trajectory (see Figure 3). In the present contribution, we also address aspects of the problem of sonifying sampled representations of state-probability distributions that are complementary to those that Williamson and Murray-Smith have described, and in particular raise questions of sample-set independence and perceptual salience of parameter mappings for sonification.

Strachan and Murray-Smith [7] described explorations with gesture input for mobile devices based on the same Dynamic Movement Primitives models that are adopted here. However, the main focus of the present contribution relates to issues of closed-loop interaction that were not entirely addressed in their application. Furthermore, the gesture recognition architecture described here is somewhat richer, capable of tracking over time an arbitrary probability density representing the correspondence between a user's gesture and an array of models.

Jenkins et al [8] have described a video-based recognition system that uses movement primitives based on an ensemble of recorded exemplars (essentially templates), together with a particle filter. Their approach is similar to the forward path of the architecture presented here, except that the movement primitives in the present contribution are based on dynamical systems models, which offer the relative advantages noted above.

Particle filters were first applied to the problem of the recognition of gestures observed in video sequences using static temporal templates in work by Black and Jepson [9]. A wide range of particle filter based methods for gesture recognition have built on their original work.

Many other gesture recognition techniques are discussed in the

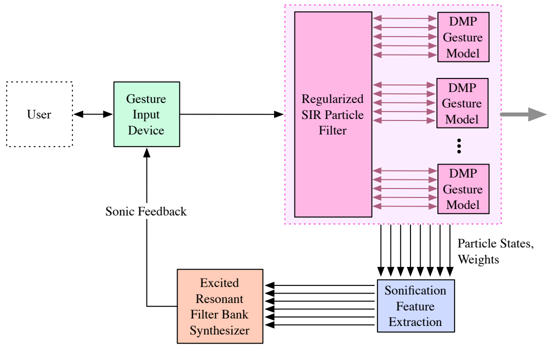

Figure 1: Architecture for gesture-based control, integrating dynamic movement primitive (DMP) models of the encoded gestures, with a particle-filter based inference mechanism, and an interactive sonification.

literature, including hidden Markov models (HMM), time-delay and recurrent neural networks, and dynamic programming algorithms, among others. Distinct advantages of the gesture modeling method we have adopted include its fine time-granularity, which is desirable in a closed loop interactive system, and its robustness to variations in the specific path an instance of a gesture may take. The fact that the system presented here incorporates generative models for the gestures provides, in addition, an important check as to whether the employed model has captured the phenomena it aimed to encode.

Further comparison of sonification methods invoked here with those used in past literature is provided in section 3.

2. GESTURE INPUT ARCHITECTURE

An overview of the system is shown in Figure 1. Gesture-based input from the user is acquired by the input device's sensors as an N-dimensional temporal trajectory. A set of particles, given by weighted states of the nonlinear dynamical systems, is evolved in tandem with the gestural input. They maintain a probability distribution that tracks the system's belief as to the correspondence between the user's input and learned models. The error signals and weights that result are then used to derive parameters for a sonification. These stages are described in more detail below, and the detailed algorithm is illustrated in Figure 2.

2.1. Nonlinear Dynamical Systems Models

The Dynamic Movement Primitives (DMP) models of Schaal et al [1][2] are adopted as nonlinear dynamical systems predicting the trajectory followed by a given gesture. An instance of a gesture is characterized by a trajectory $\mathbf{z}(t)$, consisting of a vector of values $(z_1, z_2, \ldots)$ sensed from an input device over some time interval $t \in [0, T]$. The DMP models the kinematics of the gesture trajectory with an adaptive nonlinearity f.

The system equations describing each component $z = z_i$ of a gesture are composed of a canonical system and a transformation system. The canonical system is an attractor that numerically integrates a phase variable ϕ that substitutes for the role of time:

$$\dot{\nu}/\tau = \alpha_\nu(\beta_\nu(1-\phi) - \nu), \quad \dot{\phi}/\tau = \nu \, . \tag{1}$$

The transformation system numerically integrates the dynamical variables $z_i, \dot{z}_i, \ddot{z}_i$ that describe each component of the gesture trajectory. The equations for each component are:

$$\dot{u}/\tau = \alpha_u(\beta_u(s-z) - u) + f(\phi, \nu), \tag{2}$$

$$\dot{z}/\tau = u, \quad \dot{s}/\tau = \alpha_g(g-s) \tag{3}$$

Each α and β pair represents time constants that are chosen for critical damping of the respective system, τ is a temporal scaling factor, and the goal g gives a target value for z to reach at the end of the trajectory. The term $f(\phi, \nu)$ represents a nonlinearity given the form of a weighted sum of Gaussian basis functions:

$$f(\phi, \nu) = \frac{\sum_{i=1}^{M} \psi_i \omega_i \nu}{\sum_{i=1}^{M} \psi_i}, \quad \psi_i = \exp\left(-h_i\left(\phi - c_i\right)^2\right) \, .$$

These basis functions are suitable for point-to-point gestures, possessing a velocity dependence that insures that they vanish at the endpoints of the movement. The centers c_i and bandwidths h_i are chosen so that the ψ_i are uniformly distributed over $\phi \in [0, g]$. The weights ω_i are adapted using a regression procedure such that the dynamics approximates the gesture's trajectory. Further details are provided in a related publication [10].

By contrast with finite-state based models like the HMM, where the gesture is captured by statistics associated with a sequence of discrete states, here it is encoded in the dynamical parameters of the DMP. The latter is capable of synthesizing a smooth instance of a gesture that may be varied through parameters that govern its position z, velocity $\dot{z}$, temporal scaling factor τ, and target configuration g.

Figure 3 shows several resynthesized examples of a hand-drawn figure corresponding to a horseshoe-shaped gesture. The trajectories correspond to varying the initial configuration of the DMP. Other model parameters can varied likewise. The learned figure acts as an attractor for the writing motion, which is qualitatively different from the notion of a fixed template or statistical model.

2.2. Particle Filters for Gesture Input

The particle filter tracks the system's belief over time as to the user's intended action, i.e. the correspondence of the latter with the gesture models encoded in the system. The algorithm used

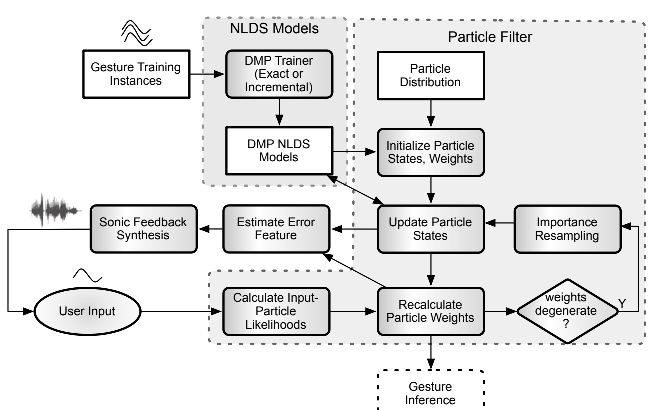

Figure 2: The gesture input, inference and sonification generation scheme.

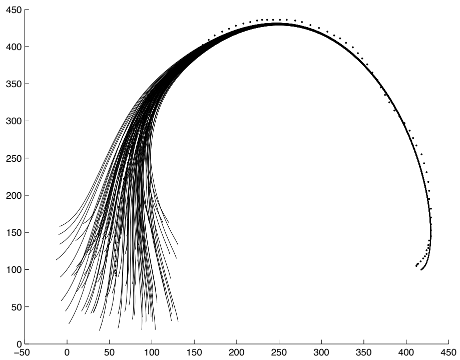

Figure 3: Several instances of a learned gesture that has been resynthesized by one of the DMP, starting from an ensemble of different initial conditions.

is the regularized sequential importance resampling (SIR) particle filter [11]. N weighted samples are employed to approximate the input gesture over time. The samples consist of states $x_t = (c, \mathbf{y}_t, \dot{\mathbf{y}}_t, g_t, \tau_t)$ characterized by:

- c A class index, indicating the DMP model whose state is being tracked
- $\mathbf{y}_t$ A vector quantity providing the instantaneous value of the gesture being modeled at time t
- $\dot{\mathbf{y}}_t$ A vector giving the instantaneous velocity of the gesture
- g_t The current target (destination posture) for the gesture
- τ_t The current temporal scaling factor for the gesture, relative to 1

At each time step, the weight for each particle is determined by a likelihood function giving the probability that the observed input trajectory $Z_t = \{\mathbf{z}_i, i = 1, \ldots, t\}$ up to time t corresponds to the trajectory $X_t = \{x_i, i = 1 \ldots t\}$ followed by that particle. We adopt a likelihood function generalized from that presented by Black and Jepson in [9], given by

$$p(Z_t|X_{t,k}) = \frac{\mathcal{Z}^{-1}}{\det^{1/2}\sigma_k} \exp(-\beta D(\mathbf{z}_i, \mathbf{y}_{t,k})) \tag{4}$$

k is the index of the k^{th} particle, β is a hand-tuned parameter governing the correlation length of the fitness function, and σ_k is an estimate of the prior variance of $\mathbf{y}$ for the class of particle k. A diagonal covariance is estimated from training instances. D is the ℓ_2 distance between the particle trajectory and the input on the given time window:

$$D(\mathbf{z}_i, \mathbf{y}_{t,k}) = \frac{1}{N} \sum_{i=t-n}^{t} (\mathbf{z}_i - \mathbf{y}_{i,k})^T \sigma_k (\mathbf{z}_i - \mathbf{y}_{i,k})) \,. \tag{5}$$

The normalizing factor $\mathcal{Z}$ in (4) is the sum over unnormalized fitness functions for all particles,

$$\mathcal{Z} = \sum_k \frac{1}{\det^{1/2}\sigma_k} \exp(-\beta D(\mathbf{z}_i, \mathbf{y}_{t,k})) \,. \tag{6}$$

The tracking process for each particle at each time step proceeds as follows:

1. Update the parameters $\mathbf{y}_t, \dot{\mathbf{y}}_t$ by integrating one time step of the corresponding DMP
2. Update the weights $w_t = p(Z_t|X_{t,k})$ using the fitness function of eq. (4)
3. If the effective particle count $N_{\text{eff}} = (\sum_k w_k^2)$ has fallen below a threshold, perform an *importance resampling* of the probability distribution, drawing N_p new particles from it to replace the old. m copies of the jth particle are created if it is drawn m times. A regularization step is used to optimally perturb the state parameters, so particles that are duplicated have nonidentical states [11].

The result is a set of particle trajectories that track hypothesis states, together with a set of evolving weights indicative of the fitness of each of the particle trajectories in describing the input gesture. The trajectories are discontinuous, due to the resampling stage.

2.3. Recognition

While gesture recognition is not the focus of this paper, the framework delineated above allows one to infer the probability that each gesture is being performed, and these probabilities may be computed in terms of the fitness functions $p(Z_t|X_{t,i})$. Assuming each of the N gesture classes to be a priori equally likely, the probability of a class c given the observed gesture Z_t to time t may be obtained from Bayes' theorem:

$$p(c|Z_t) = \frac{p(Z_t|c)p(c)}{p(Z_t)} = \frac{p(Z_t|c)}{Np(Z_t)} = \frac{\sum_{k\in c} p(Z_t|X_{t,k})}{Np(Z_t)} \,. \tag{7}$$

Where the last sum is over particles having the class label c. This equation determines the class probability given the input, within an overall factor $p(Z_t)$ that is independent of the class. A gesture can be deemed to have been recognized if this probability is greater than a given threshold and the phase variable ϕ is near to the maximum phase value of 1.

3. SONIFICATION FOR GESTURE INFERENCE

The output of the gesture input system described above consists of a set of weights $w_{t,k}$ and states $x_{t,k}$ at each time instant t. Together these furnish a sampled representation of the time-varying probability distribution $p(Z_t|X_t)$ that tracks the system's understanding of the gesture of the user.

Broadly stated, the role for sonification in this system is to convey information in order that users may better guide their actions relative to the inferences made by the system. The approach adopted hre is that of parameter mapping [12], which extracts meaningful values from the system state in order to drive the parameters of a sound synthesizer. The current situation, in which the output of an interactive system is fed to a sound synthesis engine, is notably analogous to the mapping problem encountered in the control of sound synthesis with digital musical instruments [13]. A key problem in parameter mapping approaches to sonification is the determination of salient parameters for display, which is analogous to the control mapping problem in digital musical instrument design. We describe criteria we have adopted for doing so, indicate their use in a synthesis example, and compare with alternatives that others have used in similar systems.

3.1. Considerations for Parameter Mapping

Considerations relevant for sonification in the present system include the fact that the interaction is continuous in time, that the

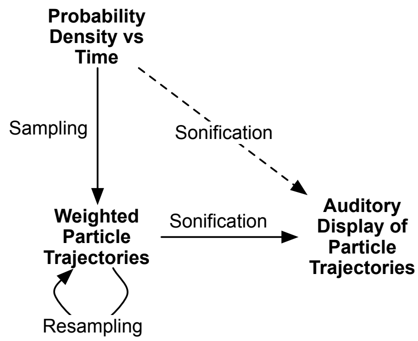

Figure 4: We suggest that a useful constraint on the sonification of sampled probabilistic models is that the sampled sonification induces a sonification of the probabilistic model which, in an appropriate limit, does not depend on the particular sample set, but only on the underlying distribution.

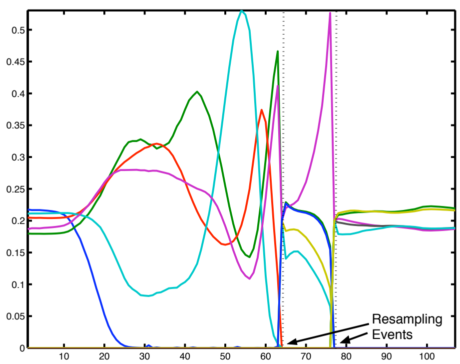

Figure 5: The signature of a resampling event in the recognizer, in which the evolving weights (shown here) and states of the particles (not shown) change discontinuously whenever a new sample set is drawn from the probability distribution.

states are composed of both discrete (class indices) and continuous (configuration) variables, and that both continuous and discontinuous changes in the time series of values $w_{t,k}, x_{t,k}$ are present. Discontinuities are the result of resampling steps, as illustrated in Figure 5, at which the probability distribution is unchanged while the sample set is redrawn. The latter changes are arbitrary, so that one wants to choose parameters and synthesis methods such that the sonification is independent of the choice of sample set.

One way to achieve this would be to ensure that parameters are extracted from sample-set invariant features of the probability distribution. All meaningful information contained in the system is accessible through the expectation values of functions of the state variables, which can in principle supply as detailed a description of the system state as desired. For example, one could adopt the instantaneous expected squared error in the estimate of the input trajectory, which is given by

$$E\{||\mathbf{y}-\mathbf{z}|||^2\} = \sum_{k=1}^{N_p} w_{t,k}||\mathbf{y}_{t,k}-\mathbf{z}_t||^2 \qquad (8)$$

As the number of particles N_p grows, the expectation value can be estimated with increasing accuracy, independent of the particular sample set. The ability to compute such quantities is a key

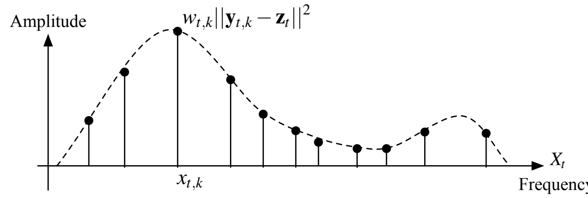

Figure 6: Example of a spectral synthesis display for the expected squared error of the gesture trajectory.

advantage of Monte Carlo probabilistic methods like the particle filter.

A better way to achieve sample set independence – and one that may intrinsically preserve information – is to present the ensemble of contributions of the various particles in such a way that the auditory display preserves features such as expectation values as perceptually identifiable aggregate properties of the synthesized sound, such as its spectral envelope or other psychoacoustic features. As an example, rather than presenting the expected value of equation (8), one could display the ensemble of weighted particle errors, given by

$$\xi_{t,k} = w_{t,k}||\mathbf{y}_{t,k}-\mathbf{z}_t||^2, \quad k = 1, \ldots, N\,. \qquad (9)$$

A mapping of the state configuration trajectory $\mathbf{y}_{t,k}$ onto the frequency of sinusoidal components of a spectral synthesis model, and of the weighted particle squared errors $\xi_{t,k}$ onto the amplitudes, would present the quantity (8), for example, as the area under under the curve given by the sound's spectral envelope at time t. See Figure 6. In such a case, the display (here, the spectral envelope) contains more information than the area alone, so with a judiciously chosen mapping, the presentation of features $\xi_{t,k}$ is more informative.

As a further example for the display of the system state during interaction, one can consider a feature derived from a measure of relative jerk between the prediction and user input. Jerk is the time derivative of acceleration. This is a quantity that the human sensorimotor system attempts to minimize during movement planning, essentially to ensure smooth movements [14]. Such a quantity may be given for each particle k at each time t by

$$J_t = ||\frac{d^3}{dt^3}(\mathbf{y}_{t,k}-\mathbf{z}_t)||^2 \qquad (10)$$

It might be employed to reinforce the idea that users' movements smoothly track the evolving state of the model over time. As the third derivative of a numerical quantity is by nature highly susceptible to noise, filtering is required to smooth the values that result. (We apply a simple rectangular window based smoothing operator.) A set of weighted particle jerks corresponding to the gestures shown in this section are displayed in Figure 7.

Other features of the probability density can be preserved with similar mappings. The display of probability densities themselves, via the weights $w_{t,k}$, may be significant toward revealing the degree of ambiguity with which the user's input is being interpreted, while the display of error-based functions provides a feedback that assesses more directly the difference between the expected and actual trajectory of sensed configurations of the input device. A range of sonification features mentioned here and elsewhere in the literature are listed in Table 1.

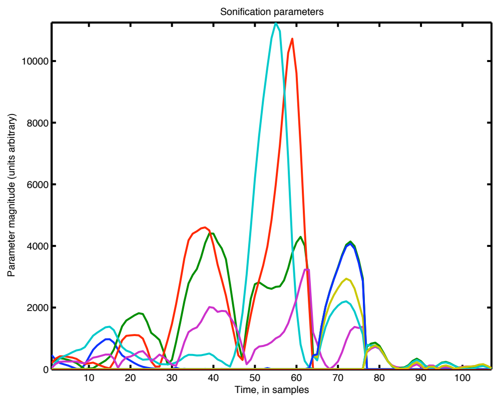

Figure 7: Examples time series of sonification features given by weighted particle jerks $w_{t,k}\mathbf{J}_{t,k}$. Those plotted here correspond to a subset of particles tracking a single gesture class for a user's handwritten figure. Time runs along the horizontal axis.

An improved understanding of the utility of the possible additional dimensions of feedback for improving control performance is an appropriate goal for research in this area. Another is to determine which features may be most useful in allowing users to perceive an array of affordances expressed as co-evolving states to that driven by the user, as in a system like that described here.

3.1.1. Sound Synthesis Method

The main focus of the present contribution is not sound synthesis methods. However, we have implemented synthesis models to test this framework. In one of these, the particle features described above are mapped onto an ensemble of time-varying resonances. For each particle trajectory, we make the correspondence shown in Table 2. The resonator model is excited with a noisy residual, and as a result, the jerk is mapped onto the noisyness of the given partial. In this way, as the user performs a gesture, the resonances of the model track the state, but become more noisy when the user's movement relative to the model flow is less smooth.

As an example, in an instantiation of the system with 1000 particles, the total number of resonances is nominally 1000. If the application demands it, because most of the particle weights are small, a threshold can be used to reduce the number of active resonances to a small enough value.

Figure 8 shows waveform and spectrogram images of the resulting sonification for a user's input gesture.

3.1.2. Limitations

In addition to the state configuration $\mathbf{y}_{t,k}$, it is desirable to encode the class c_k in the sonification, so that the jerk attributable

System parameter	**Synthesis Parameter**
$J_{t,k}$	bandwidth of resonance (α, k) at time t
$w_{t,k}$	amplitude of resonance (α, k) at t
$\|\|\mathbf{y}_{t,k}\|\|$	frequency of resonance (α, k) at t

Table 2: Sound synthesis parameter correspondences for our sonification method. In each case, the input parameters are scaled and shifted to map onto the desired range of the output parameter.

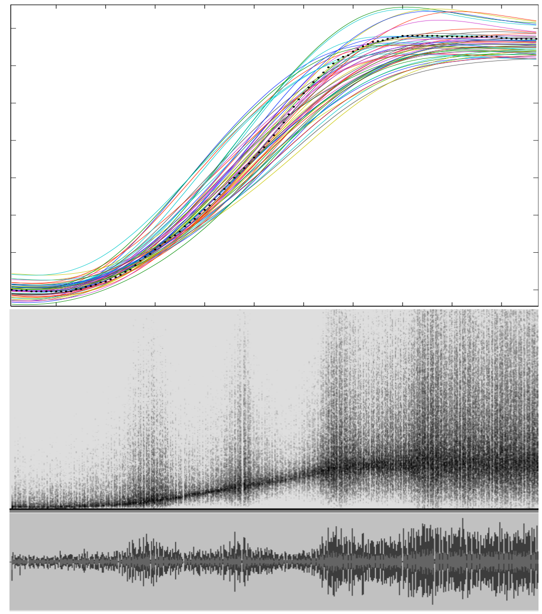

Figure 8: Waveform and spectrogram images of the sonification generated in response to the user's input shown at top. User's input can be discerned as dotted black curve in top image. Both correspond to particles constrained to a single gesture class.

to different dynamical models might be sonified in distinct ways even as the sonifications relative to distinct gesture classes are generated simultaneously. The synthesis method used here does not necessarily suggest a canonical parameter mapping that would accomplish this. However, due to the continuity of the acoustic profile of the trajectories (for example, in frequency, amplitude, and bandwidth), the user's own perception may, under appropriate conditions, be relied upon to group sounds originating from distinct classes appropriately into distinct scenes, via gestalt mechanisms [15]. We have experimented with encoding the class label in the rapid rhythmic pattern of distinct excitation sources, so as to augment the perceptual coherence of the data sources themselves.

3.2. Related Methods for Interactive Sonification

The situation discussed here fits within the domain of interactive sonification [16], and while there are several potentially relevant works from the past literature on this subject, a few are more useful for comparison with our method.

Hermann and Ritter [17] have approached the sonification of static data sets by constructing virtual physical systems around them that may be excited by a user. In one case, they describe exploring a set of potential wells through the sonification of the dynamics of an auxiliary particle system, using an audification of the kinetic energies of the particles [18]. Since we do not have the freedom to design the particle dynamics of our system, such an audification would not typically be suitable, as the data rate and system dynamics are not, in general, appropriate for generating of a sound signal with a perceptually relevant range of frequencies or temporal dynamics. This is why we have used a parameter mapping approach.

Williamson and Murray-Smith discuss in detail the utility of granular synthesis for the display of probabilistic feedback, and re-

System parameter	Name	Possible display role
$w_{t,k}$	Particle weight (conditional probability)	Display ambiguity among hypotheses Display the proximity of specific alternatives
$\mathbf{y}_{t,k}$	Particle state	Display the state of a hypothesis about the input
$\|\|\mathbf{y}_{t,k} - \mathbf{z}_t\|\|$	Particle error	Display the error of the input relative to a hypothesis Display a class or ensemble error
$d^m(\cdot)/dt^m$	mth derivative of any of the parameters above	Display the current trend Quicken the display (discussed in [6])
$J_{t,k} = \|\|\frac{d^3}{dt^3}(\mathbf{y}_{t,k} - \mathbf{z}_t)\|\|^2$	Relative jerk	Display input smoothness relative to model control policy

Table 1: Candidate parameters for the display of the particle-based gesture input system.

lated issues associated to improving gestural control over interactive systems [6][5]. The state of each particle in the sample-based displays for the control systems they describe are mapped onto the playback parameters for sound in a granular synthesis scheme. Provided a sufficient density of particles, it is plausible that the sonic texture that results evidences the underlying probability density that the particles sample, rather than the particular sample set, in the same manner as described above, and, in any event, what doesn't survive might be perceived as an unbiased noise. In any event, the situation discussed in the present contribution, in which a mutually evolving state is displayed relative to a set of models, is distinct.

4. ASSESSMENT

While we have not yet gathered results to assess the utility or usability of the system described here, trial experiments are being conducted to assess its viability, including a joint spatial and sonic figure reproduction task, using a capable nonspecialized wireless pointing device (the Nintendo Wii video game controller). We also plan to assess the system on a similar task that asks users to learn to handwrite new figures with sonic feedback. We intend to compare with results on similar tasks from the literature on haptic virtual environments, including force-feedback assisted writing [19].

5. CONCLUSIONS

We have presented a new architecture for the interactive sonification of gestural affordances, based on a combination of modern tools for gesture tracking, modeling, and new proposals for the derivation of sonification parameters.

The subject of gesture input has been of considerable interest in the human computer interaction research community for the past two decades. Arguably, the most inspirational and convincing examples for such research have come from gestural control of sound in the musical domain, comprising both mechano-acoustic control and new computer music instruments. It is thus appropriate that new approaches to this problem should return to the setting of closed loop control with sonic feedback. Together these comprise a sufficiently rich body of examples from which we hope to continue to profit as sources of inspiration.

6. REFERENCES

[1] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Proc. Royal Soc. London B*, 2003.

[2] S. Schaal, J. Peters, J. Nakanishi, and A. Ijspeert, "Learning movement primitives," *Springer Tracts in Advanced Robotics*, 2004.

[3] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W. T. Fitch, M. Grhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham ad M. Simoni, and S. Tipei, "Sonification report: Status of the field and research agenda," *Prepared for the National Science Foundation by members of the International Community for Auditory Display*, 1999.

[4] Stephen Brewster, "Overcoming the lack of screen space on mobile computers," *Personal Ubiquitous Comput.*, vol. 6, no. 3, pp. 188–205, 2002.

[5] J. Williamson, *Continuous Uncertain Interaction*, Ph.D. thesis, University of Glasgow: Glasgow, 2006.

[6] J. Williamson and R. Murray-Smith, "Sonification of probabilistic feedback through granular synthesis," *IEEE Multimedia*, vol. 12, no. 2, pp. 45–52, 2005.

[7] Steven Strachan, Roderick Murray-Smith, Ian Oakley, and Jussi Angesleva, "Dynamic primitives for gestural interaction," *Springer LNCS*, , no. 3160, pp. 325–330, 2004.

[8] O.C. Jenkins, G. González, and M.M. Loper, "Monocular virtual trajectory estimation with dynamical primitives," in *AAAI 2006 Cognitive Robotics Workshop*, Boston, MA, USA, Jul 2006.

[9] M. J. Black and A. D. Jepson, "Recognizing temporal trajectories using the condensation algorithm," in *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, Washington, DC, USA, 1998, p. 16, IEEE Computer Society.

[10] Y. Visell and J. Cooperstock, "Enabling gestural interaction by means of tracking dynamical systems models and assistive feedback," in *IEEE Intl. Conf. on Systems, Man, and Cybernetics (Submitted)*, Montreal, 2007.

[11] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2002.

[12] C. Scaletti, "Sound synthesis algorithms for auditory data representation," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed., Reading, MA, 1994, Addison Wesley.

[13] Andy Hunt and Marcelo M. Wanderley, "Mapping performer parameters to synthesis engines," *Org. Sound*, vol. 7, no. 2, pp. 97–108, 2002.

[14] T. Flash and H. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *J. of Neuroscience*, vol. 5, pp. 1688–1703, 1985.

[15] Albert S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, The MIT Press, 1994.

[16] Thomas Hermann and Andy Hunt, "Guest editors' introduction: An introduction to interactive sonification," *IEEE MultiMedia*, vol. 12, no. 2, pp. 20–24, 2005.

[17] Thomas Hermann and Helge Ritter, "Listen to your data: Model-based sonification for data analysis," in *Advances in intelligent computing and multimedia systems*, G. E. Lasker, Ed., Baden-Baden, Germany, 8 1999, pp. 189–194, Int. Inst. for Advanced Studies in System research and cybernetics.

[18] G. Kramer, Ed., *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Addison Wesley, Reading, MA, 1994.

[19] Carlo Alberto Avizzano, Jorge Solis, and Massimo Bergamasco, "Teaching to write japanese characters using a haptic interface," in *HAPTICS '02: Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Washington, DC, USA, 2002, p. 255, IEEE Computer Society.