Title: **Granular Synthesis for Display of Time-Varying Probability Densities**

Authors:
John Williamson
University of Glasgow,
Glasgow,
G12 8QQ,
Scotland,
UK
Tel: +441413303339
Fax: +441413304913
jhw@dcs.gla.ac.uk

Roderick Murray-Smith
University of Glasgow,
Glasgow,
G12 8QQ,
Scotland,
UK
and
Hamilton Institute,
National University of Ireland,
Maynooth,
Co. Kildare,
Ireland
Tel: +441413304984
Fax: +441413304913
rod@dcs.gla.ac.uk

Granular Synthesis for Display of Time-Varying Probability Densities

John Williamson Roderick Murray-Smith

***Abstract*— We present a method for displaying time-varying probabilistic information to users using asynchronous granular synthesis. We extend the basic synthesis technique to include distribution over waveform source, spatial position, pitch, and time inside waveforms. To enhance the synthesis in interactive contexts, we "quicken" the display by integrating predictions of user behaviour into the sonification. This includes linear predictions in goal space, and more sophisticated prediction using Monte Carlo sampling to predict future user states in nonlinear dynamic systems. These techniques can be used to improve user performance in continuous control systems and in the interactive exploration of high dimensional spaces. The method provides feedback from users' potential goals, and their progress toward achieving them; modulating the feedback with prediction can help shape users actions toward achieving these goals. We have applied these techniques to challenging control problems as well as to the sonification of online probabilistic gesture recognition. We are using these displays in mobile, gestural interfaces, where visual display is often impractical. The granular synthesis approach is theoretically elegant and easily applied in contexts where dynamic probabilistic displays are required.**

I. INTRODUCTION

A. Ambiguous interfaces

Human interactions with systems are difficult for two reasons: users misinterpret systems and systems misinterpret users. Solutions to the latter problem are the realm of recognition technologies and inference algorithms, but we suggest that such sophisticated interpretive mechanisms will be most fruitful if they are combined with technologies to improve human understanding of the interaction. In particular, the interaction experience can be enhanced by display of the changing state of the interaction with accurate display of uncertainty and incorporating predictive power.

The function of a human-computer interface is to interpret the actions of the user and carry out the user's intention. In practice, a system cannot interpret a user's intention with absolute certainty; all systems have some level of ambiguity. Conventionally, this is ignored; however, explicitly representing the ambiguity and feeding it back to the user can increase the quality of interaction. Mankoff *et al* describe interfaces incorporating ambiguity in [1] and [2]. This is a particularly significant issue when the system's interpretations are complex and hidden from the user. Physical buttons are intuitive because there is no hidden complexity in the interpretation. Our goal is to have systems with the all of the power of the most advanced recognition algorithms, but which remain as intuitive as a simple mechanical button. Explicit uncertainty in the interface is a step towards this goal.

Representation of ambiguity is especially significant in closed-loop continuous control situations, where the user is constantly interacting with the system to achieve some goal (see Figure 1) [3]. Formulating the ambiguity in a probabilistic framework, we consider the conditional probability density functions of sequences of actions associated with each potential goal in the system, given the current context. Examples of goals might be actions such as “Open File” or “Exit”, in a workstation interaction task. In this case, ambiguity might arise when interpreting a mouse click close to a border between items in a menu.

Users try to maximize the system’s belief about the goal they desire. The system uses its model of user behaviour to update its beliefs, and feeds back the inference so that users can act to correct any potential misinterpretation. When the probability is sufficiently high, the system can act upon its beliefs. In this context,

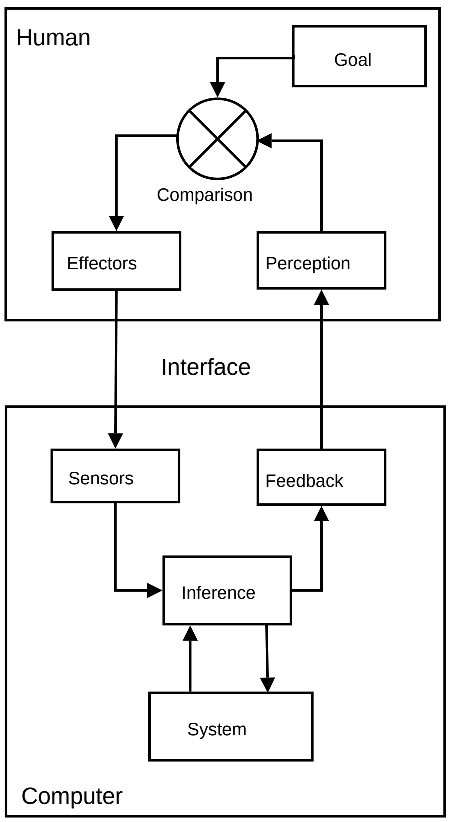

Fig. 1. A model of user interaction in a closed-loop system. In this paper we concentrate on the right hand side of the diagram, where feedback from the inference process is provided to the users which they can then compare with their goals.

providing feedback about the distribution of probabilities in the space of potential goals can assist the user, especially if the display has predictive power and can display the sensitivity of future states to current actions – “What can I do to increase the probability of my intended goal?”. Thus, we wish to sonify the time-varying properties of the distribution of goal states given potential user control actions.

B. Audio feedback

Audio can be used to present high-dimensional, dynamic data, and is suitable for use when the eyes may be occupied – for example with other visual displays, or with other tasks such as walking, when using mobile devices.

A continuous feedback model requires that the model dynamically updates the probability of each goal in real-time, and transforming this to an audio representation of the changing probabilities. At each discrete time-step t, a vector of conditional probabilities $P(\text{goal}|\text{input}) = P(G|I) = [p_1 \; p_2 \ldots p_n]$ is updated, and displayed in audio. The update can be performed by any suitable inference mechanism, so long as the probability distribution can be evaluated in real-time.

A particularly suitable method for performing this translation from probability distribution to audio is sonification via asynchronous granular synthesis.

C. Granular synthesis

1) Overview: Granular synthesis (see [4], [5], [6], [7]. For a modern, comprehensive overview of the subject see [8]) is a probabilistic sound generation method, based on drawing short (10–500ms) packets of sound, called “grains” or “granules”, from source waveforms. A large number of such packets are continuously drawn from n sources, where n is the number of elements in the probability vector.

For the discrete case, these waveforms can either be synthesized or pre-recorded. In the case of a continuous probability distribution, where n is infinite, there must be a continuous parametric form for the sources, which is generated in real-time as the grains are drawn. For example, FM synthesis could be used to represent a one-dimensional continuous distribution with the modulation index as the source parameter. After the grains are drawn, they are enveloped with a smooth window to avoid discontinuities and the grains are then summed into an output stream. Figure 2 shows the basic process.

In asynchronous granular synthesis, the grains are drawn according to a distribution giving the probability of the next grain being selected from one of the potential sources. This gives a discrete approximation to the true distribution. As the name implies, this process causes the relative timing of the grains to be uncorrelated.

The grain durations used in our implementations are of the order of 80–300ms with squared exponential envelopes; this produces a smooth, textured sound at higher grain densities. Our system generates grains such that between 100 and 1000 are always active, for a relatively accurate representation; as one grain finishes, a new one is drawn with some probability. This implies an exponential distribution on the time to generation of a new grain, and it is from this process that the asynchrononicity of the synthesis arises.

Asynchronous granular synthesis gives a smooth continuous texture, the properties of which are modified by changing the probabilities associated with each grain source. Even in situations where other synthesis techniques could be used, granular synthesis gives strong, pleasing textures which are easily manipulated in an elegant and intuitive manner. It also has the advantage that a distribution can be defined over time inside the source waveform, defining the probability of a grain being drawn from a specific time index in the wave. This allows for effective probabilistic time-stretching, which is a powerful tool in interactions where progress towards some goal is of importance (for example gesture recognition). Similar distributions over pitch (playback rate) and spatial position can also be defined.

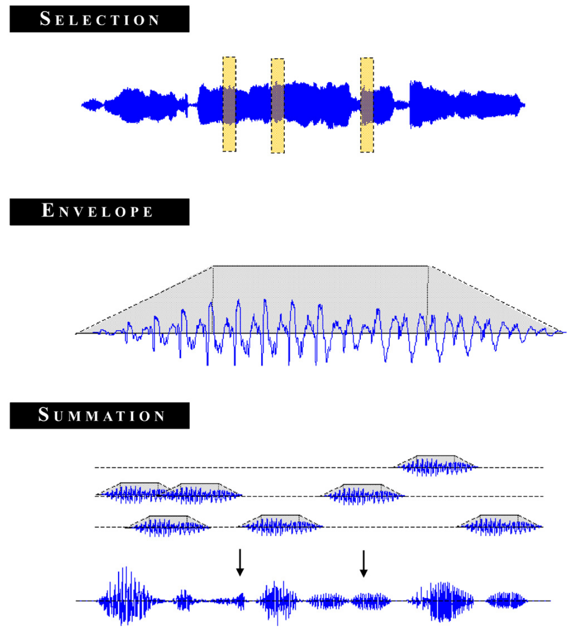

Fig. 2. Simple granular synthesis process. A much greater number of grains would be used in practice for a smoother waveform. When a new grain is created, a section of the source waveform is copied, the position of which being determined by the distribution over waveform time. This section is then enveloped. All of the currently active grains are summed to produce the final output.

D. State space representation

Figure 3 shows how a mixture of Gaussians can be used to map regions of a two-dimensional state space to sound. Each Gaussian is associated with a specific sound, and as the user navigates the space, the timbre of the sound changes appropriately. Although here the densities are in a simple spatial configuration, the technique can be applied to any state space representation by placing appropriate densities in the space. In a dynamic system the densities could be placed on significant regions such as equilibrium points or along transients. Application of these ideas to helicopter flight sonification is given in Section III-B.1.

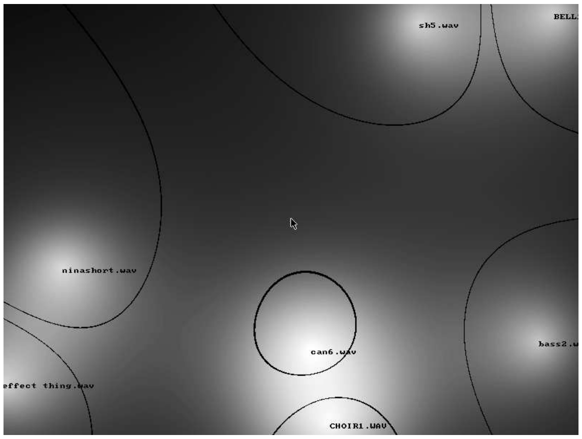

Fig. 3. Mixture of Gaussian densities in a two-dimensional state space as an illustration of the basic concept. Each Gaussian is associated with a waveform.

E. Gesture Recognition

As an example without a explicit state space representation, we extend the sonification technique to gesture recognition. A probabilistic gesture recogniser can be sonified by associating each gesture model (in our implementation, Hidden Markov Models) with a source waveform, and each model's output probability then directly maps to the probability of drawing a grain from the source corresponding to that model (see Figure 4). The temporal distribution of the grains inside the source waveforms maps to the estimate of progress through the gesture.

The design issue is then reduced to creating a suitable probability model and selecting appropriate waveforms as sources. The overall grain density remains constant throughout the sonification. In practice, this produces a sound which is incoherent when ambiguity is high, resolving to a clear, distinct sound as recognition progresses. The primary effect of the sonification is to display the entropy of the current goal distribution.

II. Display Quickening

"Quickening" (see [9], [10]) is the process of adding predictions of future states to a display. In manual control problems this allows the user to improve their performance; it has been shown that the addition of even basic predictive power can significantly improve the performance in difficult control problems:

> "Experience indicates that, by using a properly designed predictor instrument, a novice can, in ten minutes or less, learn to operate a complex and difficult control system as well as or better than even the most highly skilled operator using standard indicators" [9]

Such techniques are directly applicable to real-time sonifications of probabilistic state in interactive systems. Providing the user with information on the sensitivity of goals to inputs can allow faster and more robust exploration and control to be achieved.

In the probabilistic formulation we wish to evaluate $P(G_{t+T}|I_{1..t})$ where T is a time horizon, and t is the current time. This full distribution is usually computionally intractable, and so simplifiying approximations are made.

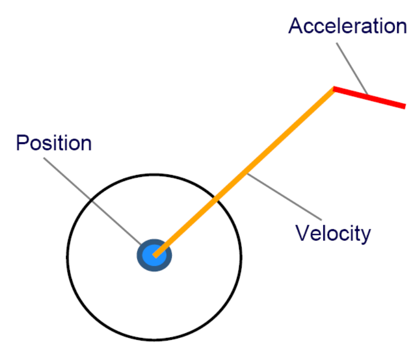

Fig. 5. A visual example of quickening. Here, the controlled object is the inner circle. The estimated velocity and acceleration vectors are shown. Similar visual displays have been suggested for use in aiding helicopter pilots in maneuvering [11].

The most basic quickening technique is the display of derivative of the variables under control (see Figure 5). In the probabilistic case, the variables are the time-varying probabilities.

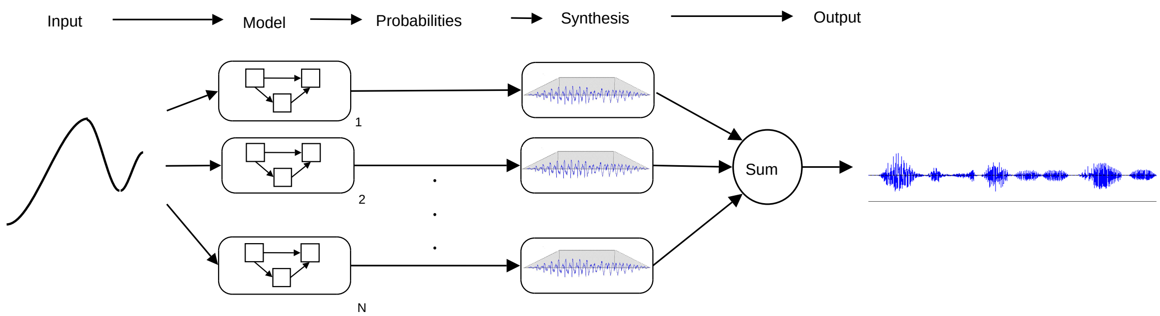

Fig. 4. Mapping from an input trajectory to an audio display via a number of gesture recognition models. Each gesture is associated with a model and the output probabilities are fed to the synthesis algorithm.

Displaying the gradient of the density along with its current value can improve the effectivity of the feedback as users can perceive whether actions are moving them towards, or away from, hypothesized goals in a continuous manner. The prediction takes place directly in the inferred space, and so is easily applicable to any problem. However, this assumes that linear predictions are meaningful in the goal space.

The granular audio display can be quickened by taking the first, second and higher derivatives of each probability p with respect to time and then forming the sum

$$v = p + \sum_{i=1}^{n} k_i \frac{dp^i}{d^i t},$$

where i is the order of the derivative and the k_i are scaling factors. v is then saturated to clip it to the range $(0, 1)$. This value can then be treated as a probability and directly sonified using the granular synthesis process described above. When the user increases the probability of a goal, the proportion of grains drawn from the source associated with this goal is increased; similarly the proportion is decreased as the goal becomes less likely. In practice, higher-order derivatives are generally less intuitive from the point of view of the user, and are also likely to be dominated by noise unless special care is taken in filtering the signals.

A. Spatial exploration example

As a simple practical example of a quickened display, the display from the exploration task in Section I-D was augmented to include first-order linear predictions. This aids users in their exploration of the space by rapidly increasing the intensity of the feedback as the users move towards the center of a goal, so that they can quickly determine which direction will give the greatest increase in probability.

In particular, the quickening is of use in accurately ascertaining the modes of the distribution. As a user approaches and then overshoots the mode, there is a rapid increase followed by an equally rapid decrease in the intensity of the feedback for that goal, allowing for faster and more accurate targeting. In higher-dimensional exploration tasks, the quickening is particularly useful for finding subtle gradients which may be difficult to perceive with an unaugmented display. As the dimension increases, increasing the weighting k_i of the derivatives can help compensate for the spreading out of the density.

III. Monte Carlo Sampling for Time-Series Prediction

Monte Carlo sampling is a common statistical method for approximating probability densities by drawing a number of discrete samples from the probability density function. This often leads to more tractable computations than directly working with the target distribution. For example, it can be used to approximate $F(p(x))$ where $p(x)$ is a (potentially complex) distribution, and F is a nonlinear function.

There is a particularly elegant link between granular synthesis and Monte Carlo sampling of probability distributions – each sample taken in the process can be directly mapped to a single grain in the output. Few other sonification methods would be suitable for directly representing the output of the Monte Carlo sampling process; approximations to the distribution would be required. Where there are more grains than samples, grains can be matched to samples in a round-robin manner. Hermann *et al*, in [12], describe a particulate approach to sonification of Markov Chain Monte Carlo for display of the properties of the sampling process. We concentrate on the use of Monte Carlo sampling for time-series prediction.

Given a model of the dynamics of a particular interactive system, where there may be both uncertainty in the current state and uncertainty in the model, Monte Carlo sampling can approximate the distribution of states at some point in the future. This can be done by drawing a number of samples around the current state, and propagating these forward according to a model of the system dynamics, with appropriate noise at each step to account for the uncertainty in the model.

A. Simple dynamic system

As a practical example of the Monte Carlo techniques, a simple dynamic system, consisting of a simulated ball-bearing rolling across a nonlinear landscape has been constructed (see Figure 7). In this system, the bearing has a state $S = [x\ \dot{x}\ \ddot{x}\ y\ \dot{y}\ \ddot{y}]$. The height component is not included in the simulation as the bearing cannot leave the surface. Here we assume Gaussian noise about the current state. The landscape model is also considered to be uncertain, in this case with a spatially varying uncertainty. In Figure 7 the dark-coloured grid on top of the lighter solid surface shows the two-standard deviation bound on the uncertainty; the uncertainty is Gaussian in this example, and so is fully specified by its mean and standard deviation.

Prediction proceeds by simulating perturbations around the current state, producing N perturbed samples. Increasing this number results in a more accurate representing of the target distribution, but at a cost of increased computational load. The model simulation is then applied to these samples for each time-step, until the process reaches $t_n = t + T$, where T is a predefined time horizon (see Figure 6 for the complete algorithm). In our example, normal ranges of the parameters are 20–40 for N and 30–80 for T. Appropriate values of the time horizon depend on the the integration constant in the simulation process and the response time of the user. Users can browse the space of future distributions by directly controlling the time horizon T. We have implemented a system with an InterTrax headtracker which allows continuous control of the time horizon with simple head movements.

In this example control actions are assumed to be constant for the duration of the prediction. Other models, such as return to zero, can easily be incorporated.

Uncertainty in the model is in this case simulated by adding Gaussian noise to the surface height at each time step, thus diffusing the samples in regions of high uncertainty. A more realistic, but computationally intensive, approach would be to draw realizations of the potential landscapes from a process with reasonable smoothness constraints, such as a Gaussian process (see [13]), drawing one realization for each of the sample paths. This would ensure that the predicted trajectories would have appropriate dynamics. The audio output proceeds as described in Section I-C.1, except that each grain maps directly to one

- Given a state $S = [s_1\ s_2\ \ldots\ s_N]$ at $t = 0$, and assuming Gaussian noise, produce

  $$a_1 = S + \mathcal{N}(0, \Sigma_s) \ldots a_N = S + \mathcal{N}(0, \Sigma_s),$$

  to get a vector

  $$A_{t=1} = [a_1\ a_2\ \ldots\ a_n],$$

  where $\mathcal{N}(\mu, \Sigma)$ denotes a normally-distributed random variable with mean μ and covariance matrix Σ^2. Here, Σ_s is the simulation noise covariance.
- Then, for each t until $t = T$, calculate

  $$A_{t+1} = f(A_t) + \mathcal{N}(0, \Sigma_m(A_t)),$$

  where f is the model function and Σ_m is the model noise.
- Each element $a_1 \ldots a_N$ of $A_{t=T}$ is then mapped to a grain.

Fig. 6. The Monte Carlo time-series prediction used in the bearing example.

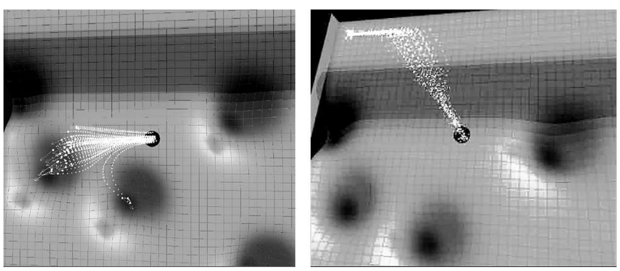

Fig. 7. Monte Carlo sampling of the distribution of future states in a simple control task. Each sample at the horizon (the arrows) corresponds to an output grain. The depressions each have a sound source associated with their mode. The darker region has high model uncertainty. On the right an example where the future trajectory passes through a region of high uncertainty is shown.

sample at the time horizon. This gives an audio display of the density at time $t = T$. This could be extended to include more sophisticated models of potential user behaviour by predicting likely future control actions and applying these as the simulation progresses. This requires a method for feeding back the control actions that lead to the final state.

B. Application domain

This display method is suitable for any continuous-control system where there is uncertainty, assuming that there also exists a reasonable model of the system which is amenable to Monte Carlo sampling. The quality of the predictions, and therefore the feedback, is completely dependent on the accuracy of the model of the system and the model of the user.

The augmentation of the display with prediction, whether in the form of complex time-series prediction or basic projection along derivatives means that latencies present in interfaces can be masked. This can be used to produce more responsive systems, and since the bound on acceptable latencies for auditory feedback is very low (around 100–200ms is the maximum before serious degradation of performance occurs [14]), this can be a very significant advantage. However, this is only feasible in cases where a reasonable model of the interface is known or can be learned. In the worst case, a poor and overconfident model of the system can lead to feedback that is an active hindrance. However, if the model makes poor mean predictions, but has an appropriate level of uncertainty in these predictions, it will still be of benefit to the user – so long as the uncertainty is displayed appropriately, as has been described in this paper. The more accurate the model becomes, the more useful the feedback will be.

Particle filters/Condensation filters (see [15]) can be sonified in an analogous manner, each of the particles in the filter being mapped to a single audio grain. Such filters are widely used in tracking and recognition tasks. For example, the particle filtering gesture recognition system described by Black *et al* [16] is ideally suited to a granular auditory display. The distribution over potential models is mapped to the distribution over waveforms, and the distributions over phase and velocity map to distributions over time inside those waveforms. Such a display completely and accurately represents the uncertainty present in the recognition system.

1) Helicopter Flight Simulation: As a concrete example of the application of these ideas in challenging control situations, we have integrated the predictive sonification into a commercial helicopter flight simulation package. We have used X-Plane[1] as the underlying engine because of its sophisticated flight models for rotary-wing aircraft and the availability of an extensive API for interfacing with the simulator in real-time.

Helicopter flight is a well-studied example of an interaction task which is known to be challenging. Control of the aircraft is difficult for several reasons: pilots must co-ordinate controls with four degrees of freedom; there is significant lag between input and aircraft response; and the aircraft is dynamically unstable (that is, it will not tend to return to a steady state and must be continuously controlled to remain stable).

There are certain states of the helicopter which are desirable (such as hovering or forward flight) which form a small subset of all the possible motions the vehicle can make. Representing the helicopter state as a point in some state space, we can define a set of goals (from the pilot's perspective) as regions in the helicopter state space. For example, the helicopter might be represented as $\mathbf{x} = [u\ v\ w\ p\ q\ r\ \phi\ \theta]^T$, where $\phi,\ \theta$ are the roll and pitch, $u,\ v,\ w$ are the linear velocities and $p,\ q,\ r$ are angular velocities. In this representation, $\mathbf{x} = 0$ corresponds to perfect level hover.

We can define *a priori* densities over such a state space corresponding to the desirable regions of flight. It is then possible to apply the previously described Monte Carlo propagation and sonification technique. A suitable approximation to the system dynamics can be obtained by taking local linearisations around the current state. This can be performed by perturbing the state of the aircraft and measuring the response, to obtain matrices A and B such that $\dot{\mathbf{x}} = A\mathbf{x} + Bu$, where u is the control input (lateral cyclic, longitudinal cyclic, collective and anti-torque, $[\theta_0\ \theta_{ls}\ \theta_{lc}\ \theta_{0t}]^T$). See [17] for a more detailed explanation of these equations. As in the ball-bearing case above, it is assumed that u remains constant throughout the prediction phase.

Setting the time horizon to around the response time of a particular aircraft produces an effective sonification of the potential future states. This is only possible because of the suitable representation of the uncertainty of the model, as the linearisation-based predictions are only accurate within a small region of the state space. Other techniques which do not represent the uncertainty would produce feedback which may lead to over-confident control and subsequent instability.

In our implementation, a number of straightforward goals were created and the Monte Carlo propagation/sonification algorithm was applied to obtain samples at the time horizon. Helicopter control tasks can be divided into a control hierarchy, so that each high-level goal is composed of some sub-goals, each of which is sonified independently. For example, "hover" is split into "level", "zero velocity" and "zero rotation". In this case, the goal sounds were chosen arbirtarily.

This demonstrates that the techniques described can easily be applied to complex systems in a very straightforward way. The implementation of this simulation example requires very little modification from the earlier example described, despite the significant complexity of the problem.

IV. Conclusions

We have presented a flexible and powerful technique for the sonification of time-varying probability distributions in the context

[1] http://www.x-plane.com/

of continuous interaction. This display represents probabilities of hypothesized user goals in a straightforward manner. In combination with predictive models to display the distribution of states at some point in the future, the interaction properties of an interface can be improved. We have applied this to both desktop interaction tasks, in the form of gesture recognition, and to nonlinear control problems. Any interactive system where some type of Monte Carlo sampling can be applied can be sonified in this manner.

REFERENCES

[1] J. Mankoff, "An architecture and interaction techniques for handling ambiguity in recognition-based input." Ph.D. dissertation, Georgia Institute of Technology, 2001.

[2] J. Mankoff, S. E. Hudson, and G. D. Abowd, "Interaction techniques for ambiguity resolution in recognition-based interfaces," in *UIST*, 2000, pp. 11–20. [Online]. Available: citeseer.nj.nec.com/321265.html

[3] G. Doherty and M. M., "Continuous interaction and human control," in *European Conference on Human Decision Making and Manual Control*, 1999.

[4] I. Xenakis, *Formalized Music: Thought and mathematics in composition*. Indiana University Press, 1971.

[5] C. Roads, "Granular synthesis of sounds," *Computer Music Journal*, vol. 2, no. 2, pp. 61–68, 1978.

[6] B. Truax, "Real-time granular synthesis with a digital signal processor," *Computer Music Journal*, vol. 12, no. 2, pp. 14–26, 1988.

[7] E. Childs, "Achorripsis: A sonification of probability distributions," in *ICAD 2002*, 2002.

[8] C. Roads, *Microsound*. MIT Press, 2002.

[9] C. R. Kelley, *Manual and Automatic Control: A Theory of Manual Control and Its Applications to Manual and to Automatic Systems*. Academic Press, 1968.

[10] R. Jagacinski and J. Flach, *Control theory for humans : quantitative approaches to modeling performance*. Mahwah, N.J.: L. Erlbaum Associates, 2003.

[11] R. A. Hess and J. G. Peter, "Design and evaluation of a cockpit display for hovering flight," *Journal of Guidance, Control and Dynamics*, vol. 13, pp. 450–457, 1989.

[12] T. Hermann, M. H. Hansen, and H. Ritter, "Sonification of markov chain monte carlo simulations," in *Proc. of 7th Int. Conf. on Auditory Display*, J. Hiipakka, N. Zacharov, and T. Takala, Eds., ICAD. Helsinki University of Technology: Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimed, 2001, pp. 208–216.

[13] C. K. I. Williams, "Prediction with Gaussian processes: From linear regression to linear prediction and beyond." in *Learning and Inference in Graphical Models*, M. I. Jordan, Ed. Kluwer, 1998, pp. 599–621.

[14] I. S. MacKenzie and C. Ware, "Lag as a deteriminant of human performance in interactive systems," in *Inter-CHI'93*, 1993, pp. 488–493.

[15] M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," *Internation Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[16] M. J. Black and A. D. Jepson, "A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions," in *European Conf. on Computer Vision, ECCV-98*, ser. LNCS-Series, H. Burkhardt and B. Neumann, Eds., vol. 1406. Freiburg, Germany: Springer-Verlag, 1998, pp. 909–924. [Online]. Available: citeseer.ist.psu.edu/black98probabilistic.html

[17] G. D. Padfield, *Helicopter Flight Dynamics: the Theory and Application of Flying Qualities and Simulation Modeling*. American Institute of Aeronautics, 1996.

Acknowledgments

Both authors are grateful for support from EPSRC *Audioclouds: three-dimensional auditory and gestural interfaces for mobile and wearable computers* GR/R98105/01. RM-S acknowledges the support of the *Multi-Agent Control* Research Training Network – EC TMR grant HPRN-CT-1999-00107. Science Foundation Ireland grant 00/PI.1/C067 and SFI BRG project *Continuous Gestural Interaction with Mobile devices*.

John Williamson received his BSc in Computing Science from the University of Glasgow, UK in 2002. He is now a research assistant and PhD student at the University of Glasgow.

Roderick Murray-Smith received degrees (BEng 90, Ph.D. 94) from the University of Strathclyde, UK. He worked at Daimler- Benz research labs in Berlin from 1990-97, was a visiting researcher in the Dept. of Brain & Cognitive Sciences at M.I.T. in 1994, a research fellow in the Dept. of Mathematical Modelling at the Technical Unviersity of Denmark from 1997-99, and since 1999 has been at Glasgow University in the Department of Computing Science. In 2001 he took up a joint position with the Hamilton Institute, NUI