Granular Synthesis for Display of Time-Varying Probability Densities

John Williamson[1] and Roderick Murray-Smith[1,2]

***Abstract*— We present a method for displaying time-varying probabilistic information to users using an asynchronous granular synthesis technique. 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 summing the derivatives of the distribution during exploration of static densities, and 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. This technique provides feedback from users potential goals, and their progress toward achieving them; modulating the feedback with quickening can help shape the users actions toward achieving these goals. We have applied these techniques to a simple nonlinear control problem as well as to the sonification of on-line probabilistic gesture recognition. We are applying these displays to 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

The function of a human-computer interface is to interpret the actions of the user to attempt to carry out the user's intention. In practice, a system cannot interpret a user's intention with absolute certainty; all systems have at least some level of ambiguity. Conventionally, this ambiguity is ignored in interactive systems; however, representing this ambiguity and feeding back to the user should increase the quality of interaction. Mankoff et al describe interfaces incorporating ambiguity in [1] and [2]. Representation of ambiguity is particularly important in closed-loop continuous control situations, where the user is constantly interacting with the systems [3] to attempt to achieve some goal (see Figure 1). 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. The goal states can be either discrete or continuous. The user's aim is then to act such that they maximize the system's belief about the goal they desire. The system uses its model of user behaviour to update its beliefs, and presents the ambiguity to the user so that they can act to correct any potential misinterpretation of their actions. Once the probability is sufficiently high, the system can act on that interpretation.

In this context, providing suitable feedback about the distribution of probabilities in the space of potential goals can assist the user. Such feedback is of particular use if the display has predictive power and can estimate and display the sensitivity of future states to current user actions – "What can I do to increase the probability of my intended goal?". We wish to sonify the time-varying properties of the distribution of goal states given potential control actions on the part of the user.

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, which require significant visual attention, when using mobile devices.

A continuous feedback model requires that the probabilistic model be able to dynamically update the probability associated with each goal in real-time, in order to produce an audio presentation of the changing probabilities. At each time-step t, a vector of conditional probabilities $P_t = [p_1 \ p_2 \dots p_n]$ is updated, and this is displayed in audio. A particularly suitable method for performing this translation is sonification via granular synthesis.

C. Granular synthesis

1) Overview: Granular synthesis (see [4], [5], [6], [7]) is a probabilistic sound generation method, based on drawing short (10–200ms) packets of sound, called "grains" or "granules", from source waveforms. A large number of such packets are continuously drawn from a n sources, where n is the number of elements in the probability vector. For the discrete case, these waveforms can either be synthesized or can be pre-recorded sounds. In the case of a continuous probability distribution, where n is infinite, there must a continuous parametric form for the source waveforms, which are then 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 then enveloped with a Gaussian window so as to avoid auditory discontinuities and the grains are then summed into an output stream.

In asynchronous granular synthesis, the grains are drawn according to some distribution giving the probability of the

[1] Department of Computing Science, University of Glasgow, Glasgow, Scotland, UK. *jhw,rod@dcs.gla.ac.uk*

[2] Hamilton Institute, National Univ. of Ireland, Maynooth, Co. Kildare, Ireland.

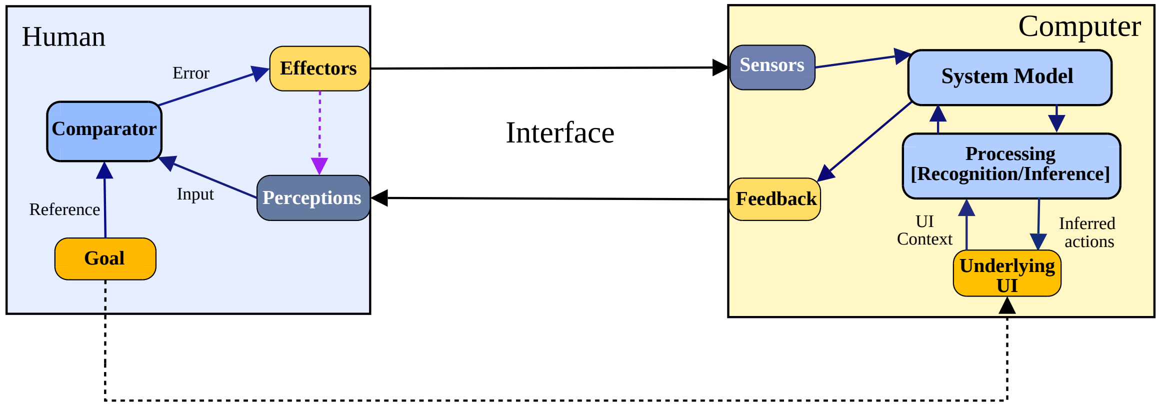

Fig. 1. A model of user interaction in a closed-loop system. The outer dashed line shows the ideal path from intention to action. The inner purple line indicates the internal control loops within the user (proprioceptive, etc.)

next grain being selected from one of the potential sources. This gives a discrete approximation to the true distribution. For values of around 50–1000 simultaneous grains this is a relatively accurate representation of the probability distribution. In our implemented system, grains are generated such that around 100–1000 are always active; as one grain finishes, a new one may with some probability be drawn. As the name implies, this causes the relative timing of the grains to be uncorrelated.

Figure 2 shows the basic process. Asynchronous granular synthesis gives a smooth continuous texture, the properties of which are modified by changing the probabilities associated with each grain source.

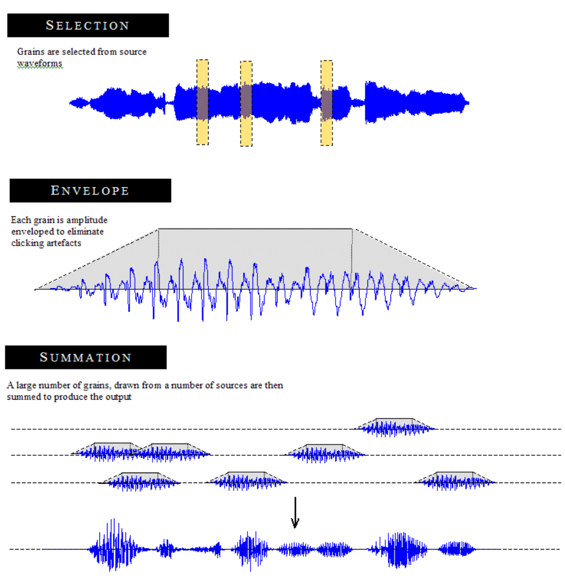

Fig. 2. Simple granular synthesis process. A much greater number of grains would be used in real output for a smoother waveform. When a new grain is created, a section of the source waveform is copied. The position of the section is determined by the temporal distribution across the waveform. This section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

Additionally, a distribution can be defined over the time inside the source waveform, giving the probability of a grain being drawn from a specific time point in the wave. This allows for effective probabilistic time-stretching. This is of use in interfaces or systems where the process of progress towards some goal is of importance (e.g gesture recognition).

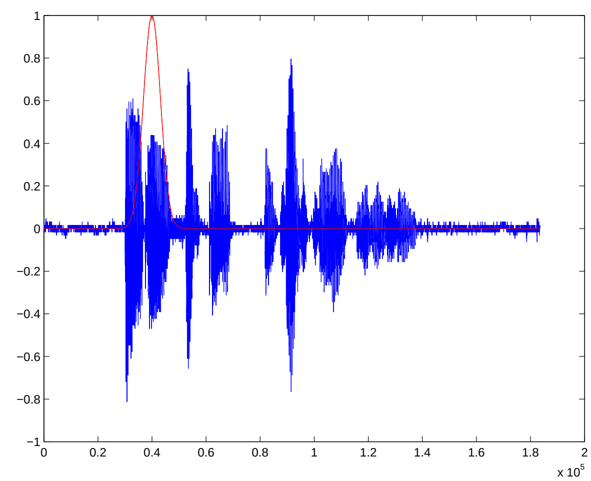

Fig. 3. Gaussian probability distribution over time in a source waveform. This can be translated from left to right to produce a time-stretching effect.

Clearly, this sound generation method is an ideal feedback mechanism for probabilistic audio display.

D. Spatial distribution example

An illustration is given in Figure 4 which shows how a mixture of Gaussians could be used to map regions of a two-dimensional state-space to sound. Each Gaussian is associated with a specific sound. 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 is general and is applicable to more abstract distributions.

E. Gesture Recognition

As an example of a more abstract system, a gesture recogniser can be sonified by associated each gesture model (such as an HMM) with a source waveform, and each model's output probability then directly maps to the probability of

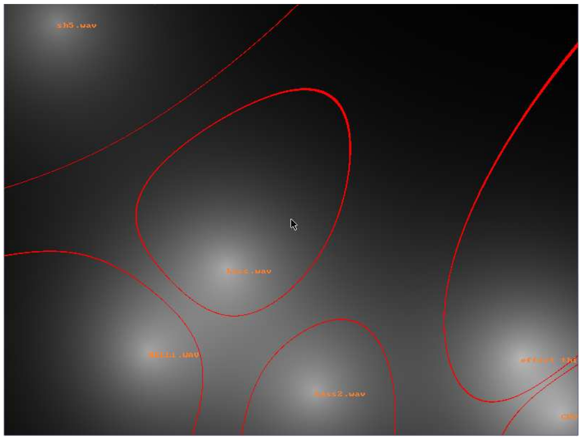

Fig. 4. Mixture of Gaussian densities in a two-dimensional state space as illustration of the basic concept. Each Gaussian is associated with a waveform.

drawing a grain from the source corresponding to that model. 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. In practice, this produces a sound which is confused and incoherent when ambiguity is high, resolving to a clear, distinct sound as recognition progresses.

II. Display Quickening

"Quickening" (see [8], [9]) is the process of adding predictions of a user's future states to the display of the current state. 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 of humans in control problems (see [8]). In a sense, the quickening is similar to the use of the integral and derivatives of the error value in automatic PID controllers.

Such techniques are directly applicable to real-time sonifications of probabilistic information in interactive systems. By providing the user with information as to the sensitivity of potential goals to the control inputs that the user may apply, faster and more robust exploration and control can be achieved.

The most basic quickening technique is the display of derivatives (see Figure 6) of the variables under control; here the variables are the time-varying probabilities. By displaying the gradient of the density, as well as its current value, the quality of feedback can be improved.

The granular audio display can be quickened by taking the 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 k_i is a scaling factor. v is then saturated to clip it to the range $(0, 1)$. This value can be treated as a probability and directly sonified with the granular synthesis process described above. The effect of this is that when the user is increasing the

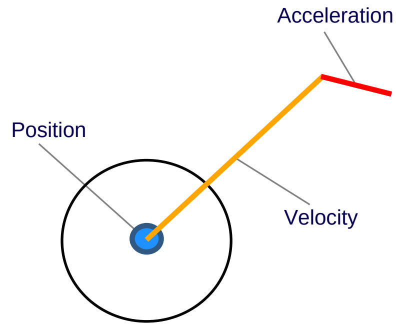

Fig. 6. A visual example of quickening. Here, the controlled object is the central circle. The estimated velocity and acceleration vectors are shown. Similar displays are used in helicopter displays to aid pilots in maneuvering (see [10]).

probability of some 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.

A. *Spatial exploration example*

As a simple practical example of the quickened display, the spatial exploration task from Section I-D was augmented with the first derivative. 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 rapidly determine which direction will give the greatest increase (or decrease) 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 a equally rapid decrease in the intensity of the feedback for that goal, allowing for faster and more accurate targeting.

In a 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 large number of discrete samples from the probability density function. This is often more tractable than directly integrating the target distribution. As the number of samples tends to infinity the samples converge to the target distribution.

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. In the case when there are more grains playing than samples taken (this may be the case if sampling is computationally expensive, and therefore the

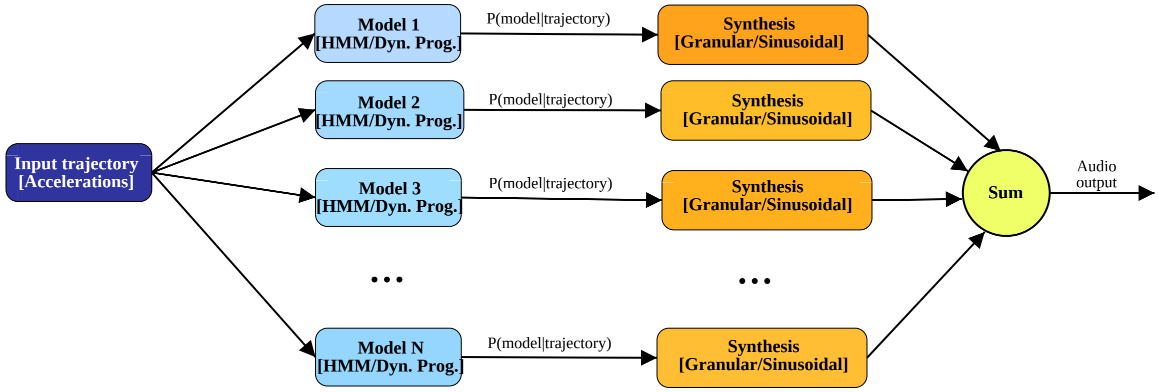

Fig. 5. Mapping from a number of gesture recognition models to an audio display

sampling is sparse) the grains can be matched to samples in a round-robin manner.

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, it is possible to use Monte-Carlo sampling to 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 the dynamics model, adding 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 non-linear landscape, has been constructed (see Figure 8). 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 is constrained so that it cannot leave the surface. There is assumed to be uncertainty in this state value; here we assume Gaussian noise. The landscape model is also considered to be uncertain, in this case with a spatially varying uncertainty. In Figure 8 the light-coloured green/red grid on top of the main solid surface shows the uncertainty (two standard deviation bound).

Prediction proceeds by simulating perturbations around the current state, producing N perturbed samples. Increasing the number of samples results in a more accurate representing of the target distribution, but at a cost of increasing the computational load. The model simulation is then applied to these samples for each time-step, until the process reaches $t = T$, where T is a predefined time horizon (see Figure 7). Note that the state perturbation is only applied at the first step, not at each of the intermediate computations. In this 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. In this example, there is no model of the future control actions of the user; control actions are assumed to be constant for the duration of the prediction.

Users can browse through the space of future distributions by allowing the time horizon T to be directly controlled. We have implemented a system with an InterTrax headtracker in

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

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

  to get a vector

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

  where Σ_s is the state 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 covariance.
- Each element $a_1 \dots a_N$ of $A_{t=T}$ is then mapped to a grain.

Fig. 7. The Monte-Carlo time-series prediction used in the bearing example.

which users can look up and down to project the predictions closer and further in time interactively.

Uncertainty in the model is in the case modelled by adding Gaussian noise to the surface height at each time step, thus diffusing the samples in regions of high uncertainty. A more realistic approach would be to draw realizations of the potential landscapes from a process which has reasonable smoothness constraints, such as a Gaussian process (see [11]), 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 now corresponds directly to one sample at the time horizon in the Monte-Carlo process. This gives an audio display of the density at time $t = T$, given that the user keeps their control actions as they are. This could be extended to include models of potential user behaviour by predicting likely future control actions and applying these as the simulation progresses. In this case a method for feeding back the particular control actions that lead to the final state is required.

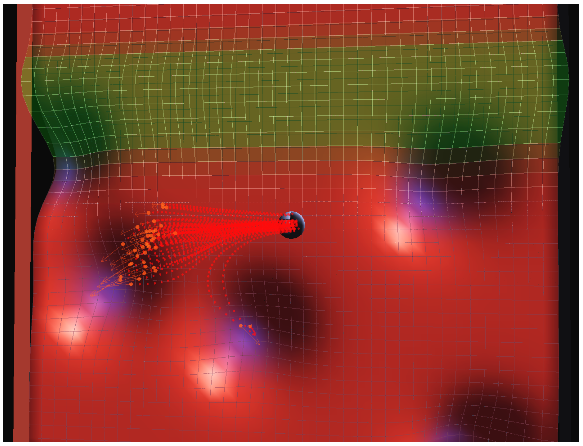

Fig. 8. Monte-Carlo sampling of the distribution of future states in a simple control task. Each sample at the horizon (the orange arrows) corresponds to an output grain. The depressions each have a sound source associated with their mode. The region in green has high model uncertainty.

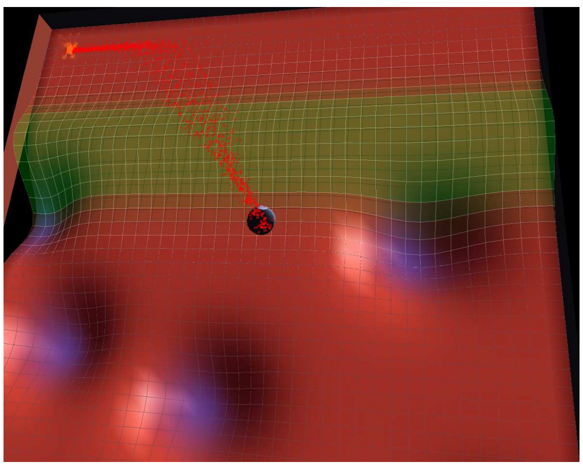

Fig. 9. Sampling where the future trajectory passes through the region of high model uncertainty.

B. Application

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.

Potential application areas of this technique include aids for helicopter control problems, where particular regions of the state-space would include hovering, stalling, or non-equilibrium transient states, for example. The pilot can then respond to this feedback with sufficient time to react. This audio presentation also gives pilots awareness of the system's sensitivity to their control actions; they can interactively explore the outcomes of control changes.

On a more basic level, we are currently augmenting a text-entry system based on continuous movement (see [12]) with the Monte-Carlo granular display to provide feedback in situations where the visual sense is already heavily loaded. In this case, we do sampling over a probability distribution given by a language model and a model of the user's dynamics. The result of the sampling is a ranked list of words (goals) and their associated probabilities.

The augmentation of the display with prediction, whether in the form of complex time-series prediction or basic projection along derivatives means that (apparent) latencies in interfaces can be reduced. 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 [13]), 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. If, however, the model makes poor mean predictions, but has an appropriate level of uncertainty in these predictions, it will still be of some benefit to the user, if the uncertainty is displayed appropriately, as described in this paper. The more accurate the model becomes, the more useful the feedback.

Particle filters/Condensation filters (see [14]) can be sonified in a similar manner. Each of the particles in the filter can be mapped to a single grain of audio. Such filters are widely used in tracking tasks and recognition task. For example, the particle filtering gesture recognition system described by Black et al [15] 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 represents the uncertainty present in the recognition system.

IV. CONCLUSIONS

We have presented a flexible and powerful technique for the sonification of time-varying probability distributions in the context of continuous human-computer interfaces. This display allows the straightforward representation of probabilities of hypothesized user goals. 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. 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. R. Kelley, *Manual and Automatic Control: A Theory of Manual Control and Its Applications to Manual and to Automatic Systems*. Academic Press, 1968.
[9] R. Jagacinski and J. Flach, *Control theory for humans : quantitative approaches to modeling performance*. Mahwah, N.J.: L. Erlbaum Associates, 2003.
[10] R. A. Hess and P. J. Gorder, "Design and evaluation of a cockpit display for hovering flight," *Journal of Guidance, Control, and Dynamics*, vol. 13, pp. 450–457, 1989.
[11] 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.
[12] J. Williamson and R. Murray-Smith, "Dynamics and probabilistic text entry," University of Glasgow, Tech. Rep. TR-2003-147, 2002.
[13] I. S. MacKenzie and C. Ware, "Lag as a deterimnant of human performance in interactive systems," in *InterCHI'93*, 1993, pp. 488–493.
[14] 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.
[15] 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

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 Enterprise Ireland BRG project *Continuous Gestural Interaction with Mobile devices*.

John Williamson 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 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 Maynooth, Ireland.