Enabling Gestural Interaction by Means of Tracking Dynamical Systems Models and Assistive Feedback

Yon Visell and Jeremy Cooperstock

Abstract— **The computational understanding of continuous human movement plays a significant role in diverse emergent applications in areas ranging from human computer interaction to physical and neuro- rehabilitation. Non-visual feedback can aid the continuous motion control tasks that such applications frequently entail. An architecture is introduced for enabling interaction with a system that furnishes a number of gestural affordances with assistive feedback. The approach combines machine learning techniques for understanding a user's gestures with a method for the display of salient features of the underlying inference process in real time. Methods used include a particle filter to track multiple hypotheses about a user's input as the latter is unfolding, together with models of the nonlinear dynamics intrinsic to the movements of interest. Non-visual feedback in this system is based on a presentation of error features derived from an estimate of the sampled time varying probability that the user's gesture corresponds to the various tracked state trajectories in the different dynamical systems. We describe applications to interactive systems for human gait analysis and rehabilitation, a domain of considerable current interest in the movement sciences and health care.**

I. INTRODUCTION

Although there have been considerable advances in the computational analysis and recognition of movement and gesture in recent years, human interaction with the systems that make use of them is frequently made more difficult than necessary. This is due to the fact that the way that gestures are being understood by the system may not be evident to users as they are enacting them, typically because users are not provided the kind of tangible and continuous feedback that is instrumental in embodying and guiding everyday interactions in the physical world. For example, a number of authors have proposed control mechanisms for intelligent homes that depend on the interpretation (by an invisible computer agent) of sign-language like motions of the hands, an approach that places a considerable burden on the motor learning capabilities of users and on the inference capabilities of the system. Problems of this nature are becoming more critical as user populations grow to include people of widely varying intrinsic sensorimotor capabilities, due to injury or other disorders, and as applications specifically targeting the rehabilitation of health related sensorimotor deficits have come to the fore [1]. Likewise, in diverse areas of product design, usability needs in the area of gesture-based interaction are becoming paramount, as spaces and devices that afford such interactions via integrated computation, sensing and actuation have grown ubiquitous. Devices already on the market range from computer peripherals to mobile phones and running shoes.

Non-visual display modalities can be preferable for improving feedback in such systems, as they may more easily be situated at an arbitrary input device, and hence closer to the locus of interaction. Further benefits of such displays include their ability to provide effective representations 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 [2][3].

A. Related work

A particle filter based recognizer for generating sonic feedback to display a user's input in applications such as a mouse gestures task has been described by Williamson and Murray-Smith [4][5]. (They describe related work on a helicopter simulator control task as well.) Their recognizer is built on 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 [6][7]. The dynamical systems models have an advantage in their ability to provide a rich prior model or the motion, representing dynamically driven variations in the detailed path taken while preserving other features of the trajectory (see Figure 2).

Strachan and Murray-Smith [8] 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 [9] 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 [10].

This work is supported in part by the Centre for Interdisciplinary Research in Music Media and Technology at McGill University.

Y. Visell and J. Cooperstock are with the Centre for Intelligent Machines and Department of Electrical and Computer Engineering, McGill University, Montréal, Canada. yon@cim.mcgill.ca jer@cim.mcgill.ca

A number of other particle filter based methods for gesture recognition have built on their original work.

Many other gesture recognition techniques are discussed in the 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.

B. Facilitating Gestural Interaction by Means of Tracking Nonlinear Dynamical Systems Models

Our system tracks the state of a user's gestures, as relayed by the sensors of some input device, and continuously estimates a correspondence between this input and a number of affordances, which are encoded as nonlinear dynamical systems models that are learned from a dataset, as described in Section II-A. A particle filter algorithm is applied as a method for tracking hypotheses as to the correspondence between the user's input and the models. The information provided by this inference mechanism is used to generate parameters suitable for non-visual display in a closed-loop interaction setting, in order to allow users to better guide their gestures relative to those afforded by the system. Pending applications of this system to gait analysis and rehabilitation are described in Section IV.

II. GESTURE INPUT ARCHITECTURE

The system presented here is illustrated in Figure 1. Input from the user is acquired by the input device's sensors as an N-dimensional temporal trajectory, and during training, these are used to adapt a set of nonlinear dynamical systems models to best match the supplied examples. During real-time input, a set of particles, given by weighted states of the learned nonlinear dynamical systems, is evolved in tandem with the gestural input. The particles 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 expected error signals relative to the system's belief, and these signals are used as parameters for a non-visual display that guides the user in continuing to enact the movement. These stages are described in more detail below.

A. Nonlinear Dynamical Systems Models

We approach the modeling problem by regarding the user's gesture as evidence of nonlinear dynamical system (NLDS) that has given rise to it. We adopt the Dynamic Movement Primitives (DMP) models of Schaal et al [6][7], which are nonlinear dynamical systems capable of reproducing the trajectory followed by a given motion. In our application, 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 $t \in [0, T]$. A DMP models the resulting kinematics by means of a set of coupled nonlinear differential equations with guaranteed convergence properties. Those employed for the system described here consist of a cascade of point attractor dynamics, with a nonlinearity that may be adapted such that the intrinsic dynamics of the DMP reproduce a given trajectory.

The system equations describing each component $z = z_i$ of a trajectory 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 $z, \dot{z}, \ddot{z}$ that describe each component of the trajectory:

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

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

In these equations, each α and β pair represents time constants that are chosen for critical damping, τ is a temporal scaling factor, and the goal g represents a target value for z to reach at the end of the trajectory. The term $f(\phi, \nu)$ represents a nonlinearity that is adapted to reproduce the desired dynamics. It is given the form of a set of weighted Gaussian basis functions

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

designed so that it vanishes at the endpoints of the movement. The Gaussian basis functions have centers c_i and bandwidths h_i which are chosen so that the ψ_i are uniformly distributed over the interval $\phi \in [0, 1]$. The weights ω_i are learned such that the dynamics approximates the gesture's trajectory, by first integrating the canonical system, and substituting the gesture trajectory into the transformation system, and solving for f. Once this is done, one can apply standard regression

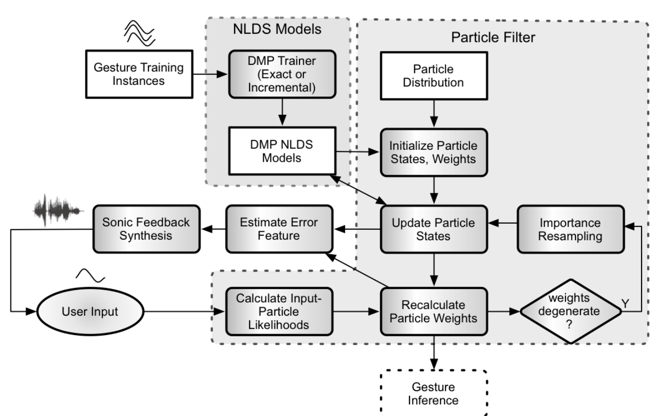

Fig. 1. The gesture input, inference and sonification generation scheme.

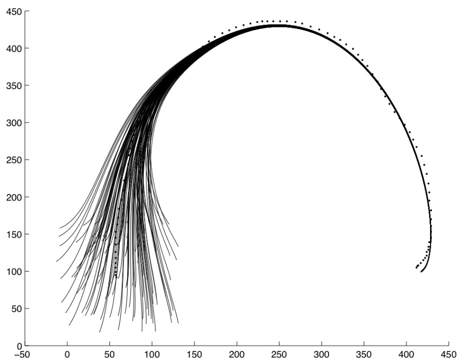

Fig. 2. Several instances of a learned gesture that has been resynthesized by one of the DMP, starting from an ensemble of different initial conditions.

formulae to estimate the ω_i from either a single gesture, or from multiple examples [6][7].

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

Figure 2 shows several resynthesized examples of a hand-drawn figure corresponding to a horseshoe-shaped gesture, as learned by a single DMP. The different trajectories, which converge over time, correspond to varying a single parameter of the model – the initial position. Other model parameters can varied likewise. The learned figure acts in this respect as an attractor for the writing motion. This is qualitatively quite different from the notion of a template or a statistical model for the movement.

B. Particle Filters for Gesture Input

The particle filter acts to track a probability distribution representing the system's belief over time as to the user's intended action, i.e. the correspondence of the latter with the models encoded in the system. Our system implements what is known as a regularized sequential importance resampling (SIR) particle filter [11]. This uses N weighted samples to approximate the trajectory describing 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 the following state variables:

- 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 fitness 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 fitness function generalized from that presented by Black and Jepson in [10], given by

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

where 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 y for the class of particle k. We estimate a diagonal covariance from training instances. D is the ℓ_2 distance between the particle trajectory and the input on the given time window:

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

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

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

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 likelihood function $p(Z_t|X_{t,k})$ using the fitness function of eq. (4). The weight update is then obtained via a fusion formula, setting $w_t \propto w_{t-1} p(Z_t|X_{t,k})$. The weight at the preceding timestep is used as the prior probability at the next timestep.
3) Compute an effective particle number, given by $N_{\text{eff}} = (\sum_k w_k^2)$, for the ensemble of weighted particles. When it falls below a threshold value, perform an *importance resampling* step (Figure 3):
   a) Resample the discrete probability distribution given by the weighted set $\{x_{t,k}, w_{t,k}\}$ by drawing N_p new particles from it. m copies of the jth particle are created if it is drawn m times
   b) Perform a regularization step [11] to optimally perturb the continuous parameters characterizing the states of the particles, so that those that are duplicated have nonidentical states

The result of this process is a set of particle trajectories that track evolving hypothesis states through the DMP models, together with a set of evolving weights indicative of the fitness of each of the particle trajectories in describing the input gesture.

C. Recognition

We have not explicitly addressed recognition in the discussion so far, as it is not the focus of this paper. However, the framework delineated above allows us 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 classes to be a priori equally likely,

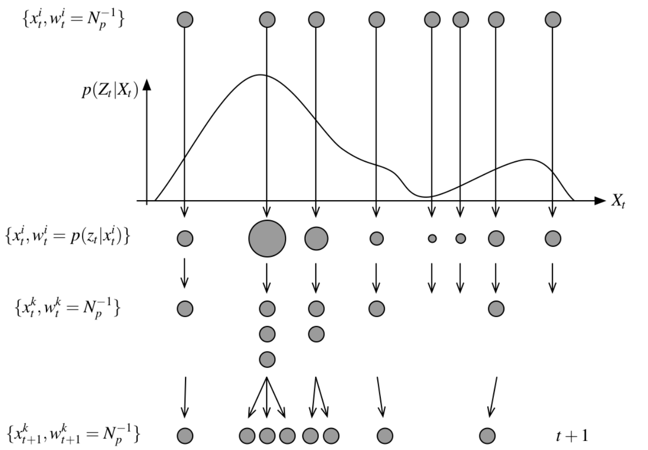

Fig. 3. The importance resampling step, schematically illustrated, sees N_p new particles with uniform weights drawn from the sampled distribution represented by the particles at timestep t. This operation is performed whenever N_{eff} falls beneath a given threshold. Those particles with the highest fitness are more likely to be copied to the next generation. An optimal regularization stage is employed to ensure that the resulting set of particles have nonidentical states.

the probability of a class c given the observed gesture Z_t to time t may be obtained from Bayes' theorem:

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

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.

III. Parameters for Non-Visual Display in Gesture Input

The output of our system 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 $p(X_t|Z_t)$ that tracks the system's understanding of the gesture of the user.

The role for non-visual display in the system is to convey information about the user's performance relative to the system's inferences as they are occurring, in order that users may better guide their actions. The approach we have taken to non-visual display generation is that of parameter mapping [12], which introduces a mapping layer to derive meaningful values from the system output for display. Criteria and considerations for parameter mapping with sampled probability densities, and an auditory display that employs them, have been described in related research [13]. A key point is to insure the perceptual relevance and coherence of the auditory display despite discontinuities in the underlying sample set owing to the resampling stage in the particle filter algorithm. For example, a display such as the spectral auditory synthesis shown in Figure 4 would be recommended based on the human auditory system's capability to perceive the spectral envelope of a sound independent of the local spectral details

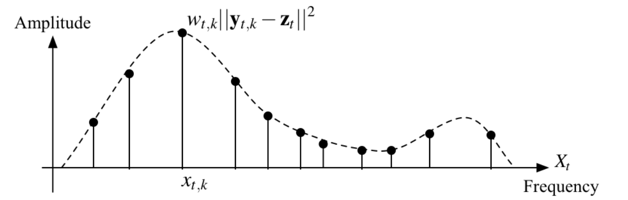

Fig. 4. Example of a spectral synthesis display for the expected squared error of the gesture trajectory.

that may articulate it. A number of parameters for non-visual display have been identified [13], and are summarized in Table I. Figure 5 shows waveform and spectrogram images of a sonification generated by this system in response to a user's input gesture.

A better understanding of the utility of the possible additional dimensions of feedback for improving movement performance is certainly one goal for research in this area (for example, Williamson [5] provides a discussion from an HCI standpoint). A complementary question that we have considered is that of determining which features may be most useful in allowing users to perceive an array of multiple afforded movements that are expressed as co-evolving states to that driven by the user, as in a system like that presented here [13].

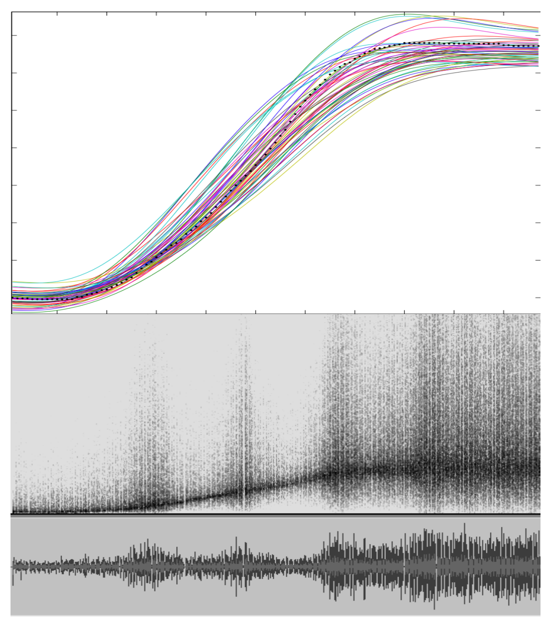

Fig. 5. Waveform and spectrogram images of the sonification generated in response to the user's input shown at top. User's input can be discerned as dotted black curve in top image. Both correspond to particles constrained to a single gesture class.

IV. Applications to the Analysis and Rehabilitation of Gait

A motivating application for the present research is the computational analysis of human walking, or gait. Recent years have seen a growth in applications of gait analysis and

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 [5])
$J_{t,k} = \|\|\frac{d^3}{dt^3}(\mathbf{y}_{t,k} - \mathbf{z}_t)\|\|^2$	Relative jerk	Display input smoothness relative to model control policy

TABLE I

CANDIDATE PARAMETERS FOR THE DISPLAY OF THE PARTICLE-BASED GESTURE INPUT SYSTEM.

in the range of techniques employed in movement sciences and clinical applications. The latter are widespread and varied, ranging from the assessment of surgery outcomes or musculoskeletal disorders affecting the lower body to neurological disorders such as strokes [14][1]. Statistical and machine learning models are becoming more broadly adopted in gait analysis. Nonlinear machine learning models (e.g. artificial neural networks) have sometimes been seen as having drawbacks that arise from difficulties in meaningfully interpreting model parameters.

Gait research has drawn extensively on the dynamical systems viewpoint to organize various concepts such as pattern generation and correlations between movement degrees of freedom [15][16][17]. This suggests an advantage of the system described in this paper over competing machine learning techniuqes, owing to the fact that DMP model parameters possess kinematic identities salient to the biomechanics of gait. Consequently, it is possible to extract meaningful gait features from learned models, or to present learned DMP model parameters in an interpretable way to therapists. Moreover, the generative nature of the model provides an additional tool through which the analyst might synthesize new gait variable time series that are typical of an individual or group, for further examination or for instructional purposes. As an example, figure 6 shows several instances of a knee flexion-extension angle time series that have been sampled from such a model learned from data acquired from 22 children. Whereas the source data constituted a digest averaged from the individual cases, a model such as that we present here may compactly represent both this average behavior and the normative range of variability in the subject group.

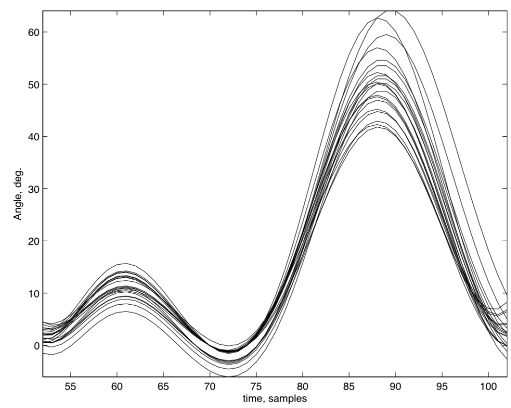

Fig. 6. Several resynthesized variants of a knee flexion-extension angle time series that was learned by a phase oscillator version of the DMP described in Section 2. (Source data courtesy of P. Selber, W. de Godoy via the Clinical Gait Analysis normative database: http://www.univie.ac.at/cga/data/index.html).

A. Non-visual feedback in gait rehabilitation

Rehabilitation of gait impairments has in the past relied on manual intervention by therapists. This is both labor intensive for therapists and of reduced benefit for patients, who can better profit if they are enabled to control their own walking movements during therapy. Benefit has recently been seen from approaches that rely on interactive systems incorporating real-time sensing, analysis and feedback via robotically supplied forces, and through visual, sonic, and vibrotactile channels [18]. In interactive rehabilitation systems, feedback typically plays a dual role of both facilitating motor control during walking, and providing motivation relative to therapist supplied goals for a training session. The neurological nature of gait impairments arising from strokes and other such disorders can increase the need for artificial feedback during therapy, as intrinsic bodily feedback channels may be impaired or absent [18]. Concrete benefits of biofeedback for gait have been found in studies with patients having a variety of disorders [19][1].

Real-time tools for gait analysis and feedback have primarily been designed according to hand crafted statistical criteria and arbitrary display mappings (e.g. from gait variable deviation or "error" to the abscissa on a time-dependent graph [18]). Detailed criteria or tools for guiding the design of visual or non-visual feedback stimuli have been lacking in the biofeedback literature, despite the rich existing literature on synthesis algorithms (e.g. sound and haptic synthesis) and on the perceptual and task salience of display mappings (e.g. auditory display) in human-computer interaction. Approaches such as that presented in this paper that are based on a fusion of analytic models with error-based feedback could prove useful toward creating more informative and usable displays for rehabilitation systems.

V. Conclusions

We have presented an architecture for enabling movement and gesture based interaction in a way that incorporates concurrent assistive, non-visual feedack, and have outlined future applications to the improvement of systems for gait analysis and rehabilitation. While recent theoretical and practical advances in the computational understanding of movement have been considerable, a significant challenge has emerged out of the observation that many of those in a position to be most aided by these technologies are also those who may have the greatest difficulties in using them. An approach to addressing this dischord may be to design systems to be able to adapt to and compensate for the sensorimotor deficits of their users, as we have attempted to do here. More concretely, they may be organized to provide for both the understanding of users' actions and the stimulation of their senses in a closed-loop negotiation of the sort that reflects the interplay between perception and action that surround everyday tasks in the real world [20]. This represents a key goal informing our ongoing research in this area.

References

[1] Michelle J. Johnson, "Recent trends in robot-assisted therapy environments to improve real-life functional performance after stroke (commentary for journal special issue)," *J. of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, 2007.

[2] 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.

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

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

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

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

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

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

[9] 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.

[10] 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.

[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] Y. Visell and J. Cooperstock, "Modeling and continuous sonification of affordances for gesture-based interfaces," in *Proc. of the International Conference on Auditory Display (submitted)*, Montréal, Canada, June 2007.

[14] R. Begg and M. Palanis, Eds., *Computational Intelligence for Movement Sciences: Neural Networks and Other Emerging Techniques*, Idea Group Publishing, Hershey, PA, 2006.

[15] S. Kelso, *Dynamic patterns: The self-organization of brain and behavior*, MIT Press, Cambridge, MA, 1995.

[16] J. Hamill, J.M. Haddad, and W.J. McDermott, "Issues in quantifying variability from a dynamical systems perspective," *J. of Applied Biomechanics*, vol. 16, pp. 409–420, 2000.

[17] R.E.A. Van Emmerik, M.T. Rosenstein, W.J. McDermott, and J. Hamill, "Nonlinear dynamical approaches to human movement," *J. of Applied Biomechanics*, 2004.

[18] R. Riener L. Lunenburger, G. Colombo, "Biofeedback for robotic gait rehabilitation," *J. of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, 2007.

[19] R. W. Teasell, S. K. Bhogal, N. C. Foley, and M. R. Speechley, "Gait retraining post stroke," *Top. Stroke Rehabil.*, vol. 10, no. 2, pp. 34–65, 2003.

[20] A. Noe, *Action in Perception*, MIT Press, Cambridge, MA, 2004.