

Pointing Without a Pointer

John Williamson¹

¹Department of Computing Science,
University of Glasgow,
Scotland
jhw@dcs.gla.ac.uk

Roderick Murray-Smith^{1,2}

²Hamilton Institute,
Nat. Univ. of Ireland Maynooth,
Co. Kildare, Ireland
rod@dcs.gla.ac.uk

ABSTRACT

We present a method for performing selection tasks based on continuous control of multiple, competing agents who try to determine the user's intentions from their control behaviour without requiring an explicit pointer. The entropy in the selection process decreases in a continuous fashion – we provide experimental evidence of selection from 500 initial targets. The approach allows adaptation over time to best make use of the multimodal communication channel between the human and the system. This general approach is well suited to mobile and wearable applications, shared displays and security conscious settings.

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces – Interaction Styles; H.5.2 [Information Interfaces and Presentation]: User Interfaces – Theory and methods

General Terms: Design; Theory

Keywords: Probabilistic, selection

Introduction

Selection without the use of pointing is of particular current interest because of challenges faced in designing interaction with mobile devices and wearable computers, where input and output are often limited, or unconventional means of interaction are available. These devices might be instrumented with sensors for which pointing may not be a natural method of selection, (e.g. accelerometers), or may have displays (such as audio displays) that are not suited to pointing. The display may also just be too small to allow convenient pointing.

In this paper we present a method for performing selection tasks without a pointer, based on a continuous interaction interface, which can be flexibly linked to a wide range of input and feedback/display mechanisms. Our method is easily generalised to controllers with many degrees of freedom, allowing us to make use of rich, multimodal input methods. It is an intrinsically probabilistic interface, which can incorporate

knowledge of uncertain prior information and noisy inputs, when adapting the dynamics of interaction with the user, and which can flexibly and continuously adapt to changes in context.

Avoiding a pointer also has a potentially interesting role in multi-user interaction with a shared display or in single user interaction where multiple independent limbs (for example, bimanual control or finger control) can perform selection simultaneously. The absence of a pointer also has the advantage that in applications where security issues are important, we can design interfaces with informative feedback which is useless to an observer who is not in the interaction loop. We may also want to perform rapid selection from very large numbers of objects; conventional visual pointer feedback makes this difficult, especially on small or unconventional displays. Such an interface is particularly suited to use with inertially sensed inputs on mobile devices, where direct pointing is often non-intuitive.

Continuous interaction and manual control

Our interface is an example of *continuous interaction*, which has been described as follows:

“... *the user is in constant and closely coupled interaction with the computing system over a period of time. The interaction is no longer based on the exchange of discrete messages that could be considered atomic actions, but the input provided by the user and/or the output provided by the computing system are a continuous process of exchange of information at a relatively high resolution...*” [1].

It is also an example of an interface built on methods from manual control theory – the study of how humans control dynamic systems – see [2] for a recent review of this field. These methods are close to those used in perceptual control theory, proposed by William Powers [5, 6]. This earlier work suggested that many kinds of behaviour can be described as continuous control problems, and Powers proposed that this viewpoint provides an empirical method for the estimation of a subject's intention. This can be done by designing an experiment to identify which variables the subject is controlling, by introducing changes (*disturbances* in control theory nomenclature), directly or otherwise, to variables which are under the subject's control. Variables for which disturbances are corrected are assumed to be controlled by the subject.

A Multi-Agent Perspective

We will now adapt this approach to develop a continuous, multimodal selection interface suitable for a wide range of applications. One way of applying this to a user interface is to consider interface components as independent agents competing for user attention. Each agent would be associated with an action in the user interface. Agents try to determine whether the user is interested in them by designing and running “experiments” to look for correlated responses in the actions of user. The experiments can take any form where the agent changes its state and tests for correlated responses in its inputs (movements of the mouse, for example). As the user’s attention is a scarce resource, the agents must compete to optimally ascertain user intentions with limited attentional resources.

The communication channel between the system and the user is bandwidth-limited both perceptually and physiologically. This bandwidth, which is time-varying, is affected by both low-level physiological responses and higher-level cognitive constraints such as available attention, stress levels or skill.

Implementation Example

As an example of this technique, a simple selection interface similar in nature to the experimental designs used by Marken [4] has been constructed. In this system, there are a number of objects which can be selected (represented visually) which are subjected to a smooth pseudo-random disturbance. Alternatively, a mixture of sinusoids (normally more than five) with frequency and phase randomly selected from a suitable frequency band can be used as a disturbance – this approach often used in manual control work (e.g. [2]) because of its theoretical predictability, but apparent perceptual unpredictability.

Movement of the mouse applies the same control action to all of the objects simultaneously. Selection is achieved by stabilising one of the objects for a period of time so that correlations are introduced between the control inputs and the agent disturbances. It is not necessary that the objects are fully stabilised, as long as correlations between the input and disturbance are present, although complete stabilisation provides stronger feedback. An under-actuated display would be useful in secure applications where it is desired that passive observers not be able to observe the final selections. As in many other continuous control problems, improvements in user control can be achieved by “quickenning” the display (see [2, 3]), for example by displaying velocity vectors in a position control system.

Viewed from an agent perspective, each object is considered to be an agent whose experiment is the disturbance; intention is detected by looking for controlling behaviour. In this way, selection can be carried out without having to select directly with a pointing instrument. The agents produce a continuous probability of selecting agent i , $p_i(s)$ or just $p(i)$. A prior distribution over likely actions can easily be introduced to reweight these probabilities.

One advantage of this system over other selection devices is its robustness to accidental selection. As the probability of the control input being correlated with the pseudo-random disturbance is very low (depending on the time-frame across which they are compared), the selection process requires intentional control and knowledge of displayed information to operate. The degree of correlation required is an adjustable design parameter, so that designers may easily create interfaces with robustness appropriate to their context of use. The statistical nature of the disturbances means that rigorous significance tests can be exploited to determine appropriate robustness parameters for different contexts. This may be of value in mobile devices, for example, where accidental activation of functions while the device is being carried may cause problems.

The objects in this example are abstract entities but could, for example, represent letters in a text-entry system or buttons in a conventional user interface. Such an interface can present a very large number of objects simultaneously without requiring accurate pointing skills which may be difficult in normal conditions of use. The information content required to select remains the same, but the time taken to perform the selection can be tuned by adapting parameters in the system.

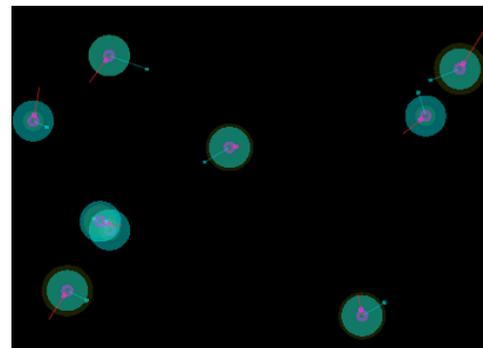


Figure 1: Prototype screenshot. Each agent is shown as a circle with radius proportional to selection probability. The velocity of the objects is shown as an aid to users.

Algorithm

The following algorithm is used as a simple example of a method for performing the selection which can be implemented in a very straightforward way. Much more sophisticated and robust methods for detecting the control behaviour of the user are possible, using methods from information theory to estimate mutual information between signals, and from manual control to incorporate expected lags induced by the human controller. Our current implementation fits a first-order lag to the user control behaviour during a calibration sequence. In this case, however, we limit ourselves to the variance-based algorithm below, which is particularly simple.

Detail Each object has a position f_i to which is added low-pass filtered white noise. The displayed position of the object is the sum of this position and the mouse position m . Com-

putting the variance of this sum over a time window T , we obtain σ_{si} . The variance of f_i alone, σ_{fi} is also computed.

To compute the probability p_i of object i being selected, an intermediate value γ_i is calculated. Given $r_i = \frac{\sigma_{si}}{\sigma_{fi}}$, a threshold v with $v < 1$ is applied, and objects with $r_i < v$ increase γ_i additively with a factor αr_i ; values below this decrease multiplicatively by a factor β . Normalizing the γ_i 's produces the output probabilities p_i . The intermediate integration/exponential decay stage stabilises the selection process. Tuning α and β adjusts the time required to select an object. Effectively, this procedure assigns higher probability to objects which have lower high-frequency spectral content over a period of time.

In our implementation of this algorithm, the selection probability is displayed to the user visually. A selection event is considered to have happened when p_i exceeds some threshold; values will depend on the speed at which selection can take place versus the risk of ambiguity.

Results

A comparison of spatial trajectories between an object which the user is intentionally controlling and an object which the user is not interested in is shown in Figure 2. The probabil-

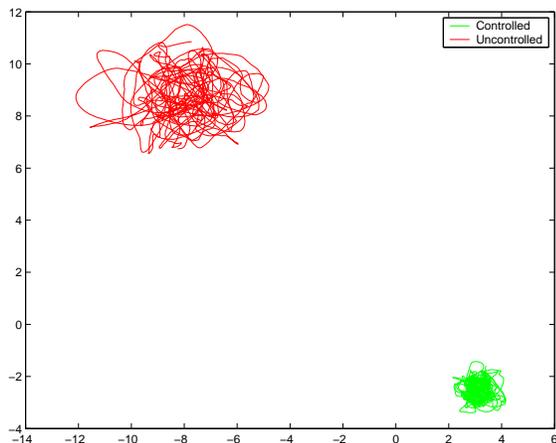


Figure 2: Trajectories of two objects, one of which is being selected. The reduced variance of this object (lower right) is visible.

ity time series produced by the agents during repeated selection tasks for different numbers of on-screen objects in the test implementation are shown in Figures 3–5. It should be noted that the entropy changes *continuously* in this interface, unlike in a menu hierarchy, say, where entropy, and any associated probabilistic feedback, is only adjusted at each discrete selection event.

User Testing

Informal testing with a number of computer-literate users was carried out as part of the prototyping. The experiment was carried on on a laptop with 15 inch screen and a mouse for input. Although sometimes users were initially confused by the selection metaphor (one commented that the display

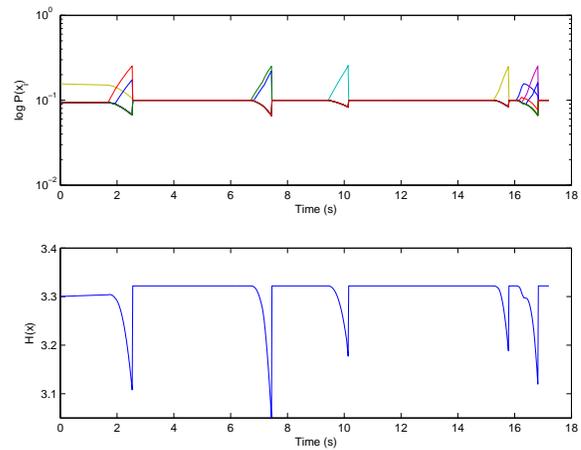


Figure 3: Probability time series for ten objects with a number of correct selection events. Entropy $\left(\sum_{i=1}^N P(x_i) \log_2 \frac{1}{P(x_i)}\right)$ is shown as the blue line below.

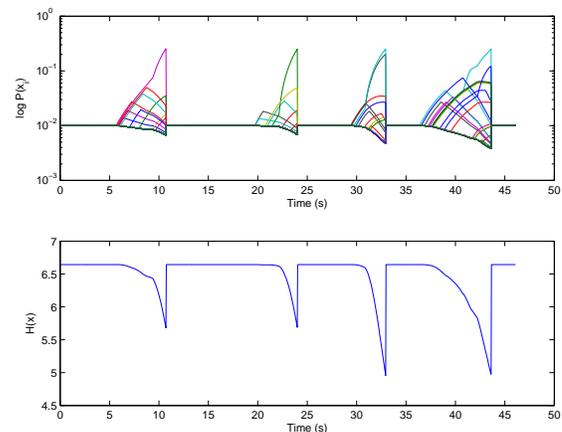


Figure 4: Time series and entropy as Figure 3, 100 objects

would be less confusing if it showed the direction to move to counteract the disturbance, rather than the direction of the disturbance), all were able to use the interface for selection after five minutes practice. Untrained users performance for selecting one object from 100 was around 20s, though more experienced users reached times of around 5–10s. Times were not significantly faster for 10 objects with most users able to select in between 4 and 10 seconds, suggesting that the interface is better suited to larger numbers of agents. The actual selection times are completely dependent on the robustness settings of the interface.

With the system implemented on an iPaq PDA, using an accelerometer for tilt control, users found the selection to be significantly more intuitive and were able to use the system almost instantly. This is a interface where traditional pointer-based selection would be clumsy.

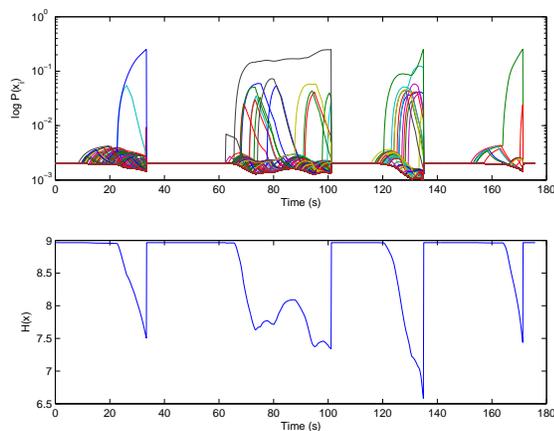


Figure 5: Time series and entropy as Figure 3, 500 objects

Alternative Disturbances

In this example, there are two degrees of freedom, since a mouse is being used as an input device. If, however, higher dimensional controllers (such as data gloves, body suits or spline controllers) are used, the technique can be extended similarly. The time taken to perform a selection will be smaller with a higher dimensional controller, as the bit-rate will be higher with greater degrees of freedom. The actual bit-rate attainable will depend on the user's ability to co-ordinate multiple channels simultaneously and independently.

Many different forms of presentation of the disturbances are possible; changes in scale, orientation, shape, colour or other attributes are all equally applicable. The agents could be laid out in a lattice formation with changes in orientation rather than position; this would mitigate the confusion that some users noted with the positional prototype. Other modalities can also be used for the presentation of the object state; for example, a one-dimensional example could produce disturbances in pitch of a sound.

Correlation Structure

To increase the value of the control inputs in disambiguating the potential agents, the disturbance patterns can dynamically attempt to provide maximum decorrelation between the most likely elements. This can, for instance, be achieved by increasing the frequency range of the disturbances as the probability of the agent increases. This allows for a slow, stable interface until selection activity occurs. The bandwidth then increases and the user can provide more information to identify the intended agent. This, combined with appropriate audio and potentially vibrotactile feedback can be used to give the user the feeling of *exciting* a target, despite the notional stabilisation of the target.

More complex correlation structures between agents in the system can be introduced. Agents can be grouped by giving them some common disturbances which are combined with individual disturbances to produce the output. This allows the selection and highlighting of the group of related items by following the common disturbance. The agents can

then be selected individually. This could be useful where objects belong to multiple classes and these groupings can be selected by controlling subcomponents of the disturbances.

Adaptive Agents and Multimodality

The agents in the interface can be designed so as to adapt over periods of time, attempting to optimize the "experiments" they run to best ascertain the user's intention. Adaption can take place through re-weighting of disturbance outputs depending on the previous success in determining user intention. If multiple feedback modalities are used, this reweighting can dynamically adapt the interface to fit the context. If, for example, a sighted user presented the system to a blind user, the visual weighting would gradually decay as the agents realize that they can obtain no information through that channel. Or, in a mobile context if the environment changes from quiet to noisy, audio feedback will be less successful in determining intentions and additional weight can then be redistributed to visual and haptic modalities.

Conclusions

We have presented a framework for probabilistic selection interfaces in continuous environments without a pointer. This is a flexible and extendable technique for utilising rich input devices and adapting to available feedback channels. The basic approach is extremely simple to implement, but there is a great deal of research opportunity in finding natural, intuitive 'disturbances' and correspondence mappings between qualitatively different display and input modalities.

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*.

REFERENCES

1. Doherty, G., and Massink, M. Continuous interaction and human control. *Proceedings of the XVIII European Annual Conference on Human Decision Making and Manual Control*. Group-D Publications, Loughborough (2000), J. Alty, Ed., 80–96.
2. Jagacinski, R. J., and Flach, J. M. *Control Theory for Humans: Quantitative approaches to modeling performance*. Lawrence Erlbaum, Mahwah, New Jersey, 2003.
3. Kelley, C. R. *Manual and Automatic Control*. John Wiley and Sons, Inc., New York, 1968.
4. Marken, R. S. *Mind Readings: Experimental Studies of Purpose*. The Control Systems Group Book, 1995.
5. Powers, W. T. *Living Control Systems: Selected papers of William T. Powers*. The Control Systems Group, 1989.
6. Powers, W. T. *Living Control Systems II: Selected papers of William T. Powers*. The Control Systems Group, 1992.