Triggering models for effortful musical interaction beyond direct cognitive control

Dr Zlatko Baracskai
University of the West of England
zlatko.baracskai@uwe.ac.ukrg

ABSTRACT

As the field of musical computer interaction embraces the need for difficulty of control and elongated learning curves we explore some possibilities in designing such instruments. This paper discusses models designed to derive triggers from sensor data putting emphasis on mapping and its potential for hosting effortful interaction design. Focus on deriving triggers entails achieving different triggers and trigger characteristics while the timing accuracy isn't getting discussed. Starting from basic arraying and thresholding principles complex schemes are devised such that the levels of difficulty are built into the system. It is proposed that these mapping strategies designed for musical performance bear resemblance to acoustical instruments in the way they accommodate control at many skill levels. The purpose of this paper is to exemplify such systems and describe their performance based on informal tests in order to demonstrate the potential of complex mapping strategies paired with simple interfaces. In appropriating interfaces to be operated at different levels of expertise it is anticipated that an investment of rehearsal time will produce increased accuracy of control, as well as the physicality of virtuosic performance, which allows novice audiences to more deeply appreciate computer facilitated instruments.

1. INTRODUCTION

Presented triggering models attempt to offer 'no ceiling' and 'low entry fee' [1] using sensor signal processing models that are generic enough to function as a control paradigm, regardless of the sound producing solutions employed or the sensor system itself. The underlying idea is that an array of sensors or sensor signals could provide higher-level information than each sensor separately, akin to three-finger description and parameterisation of a triangle [1]. This has led to an initial model that is based on four tilt switches, which will be described in more detail later. In developing such systems it soon became clear that even a single stream of sensor data can lend itself to extraction of information, which can be musically useful beyond raw data transferred to the computer system, akin to early achievements audio analysis [2]. Many basic control analysis modules have been developed over the past years that are freely available for the Max for Live platform [3]. The proposed systems and many others can be thus built using a modular approach without necessary fluency in programming. In conclusion to the conclusions inspired by the models described below, this paper proposes a distinction of two diverging approaches in designing sensor signal processing models.

2. CASE STUDIES

As the theory developed in this paper comes from practical solutions the introduction of concepts will be based on documenting three case studies. As such, this account of techniques that can be employed for the purpose defined earlier is by no means exhaustive. It is aimed to provide reference to systems that fulfil the proposed characteristics, as well as to inspire further developments of sensor signal analysis in the community.

2.1 Primitive Arm-tracking

The initial sensor signal processing model developed was based on an array of tilt-switches. These simple sensors, used widely in pinball machines, seemed challenging to employ within a dance performance interface for musical control. A tilt-switch is a rather inaccurate way of measuring tilt because the sensor reading responds to erratic movement just as the horizontal position cannot be established other than by tracking rapid switching behaviour which is obviously prone to error. This is due to a simple design of a moving ball within a closed tube that establishes contact between terminals when it touches one of the tube ends. As such, these sensors will provide a distinction between two possible inclinations if the angles are sufficiently large and the motion is steady. This property makes the control of the system challenging to the extent that initial trials might adversely affect the motivation to practice the skills necessary for accurate control, introducing real difficulty and challenge. An array of these binary switches was placed along the arm of a dancer starting at the top of the shoulder, such that the moving joints separate them. Taking all possible combinations of four binary switches one can outline the 16 distinct configurations that such an array can assume.

These are depicted in Fig. 1, where the ball represents the shoulder while the arrow represents the hand.

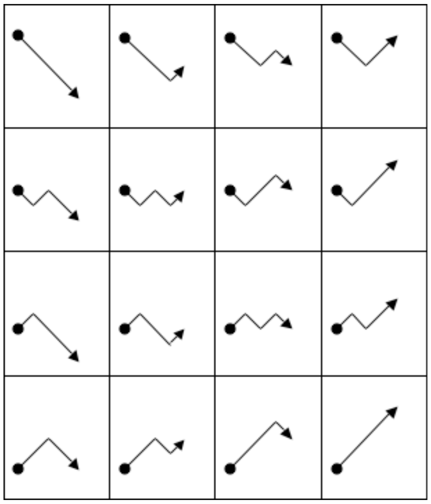

Figure 1. Sixteen states of an arm.

Assuming that the upward tilt in the direction of the arrow is represented by state 1 and the downward tilt by state 0, each of the 16 states can be matched to a 4-bit binary number. If the hand sensor supplies the least significant bit then Figure 1 representing sixteen possible configurations matches Table 2, which spells out the construction of 4-bit binaries in the same arrangement.

0000 = 0	0001 = 1	0010 = 2	0011 = 3
0100 = 4	0101 = 5	0110 = 6	0111 = 7
1000 = 8	1001 = 9	1010 = 10	1011 = 11
1100 = 12	1101 = 13	1110 = 14	1111 = 15

Table 1. Binary numbers representing 16 arm positions.

Achieving different positions in this model bears different levels of difficulty for a performer. Rather obviously, assuming positions in the top left and bottom right corners and would bear little cognitive load easy whereas black shaded positions require both mental and physical effort. It seems that the level of difficulty is proportional to the amount of bent joints along the arm, which is the number of alternating values along the bit string. As there are three connecting joints (shoulder, elbow and wrist) and thus three bit alternating possibilities, the levels of difficulty can be discretized in four values. The level of difficulty is thus represented in Table 1 by the background shade of the cells, darkest shade representing the highest level of difficulty.

This theoretically devised model proved to be quite difficult to control in performance. Considering that all positions would be reliably assumed only if none of the sensors was close to a horizontal position, it becomes clear that most of these shapes are far from an arm's natural repertoire. It is likely that a better interface could be designed with accurate arm tracking systems employing finer sensor technologies. Nevertheless, the transferability of this triggering model and the inherent differentiation of states according to their complexity makes it a valuable design in support of effortful and virtuous performance that will be later discussed. Let this serve as a basic example of combining event-type data to yield a performatively more useful set of event descriptions through **dimensional manipulation**. In this case four distinct binary axes are combined into a dimension with 16 discrete values available.

The intent of arraying sensors such that they form a larger, coherent unit could be taken further in this case, quite simply by including both arms. In this way the difficulty of achieving configurations can be further discussed and differentiated. The proposed bit alteration examination is the most straight-forward option and further models could be devised with more accurate consideration of psychology and physiology. Nevertheless, a triggering system needs certain simplicity for it to fulfil its function, which is to establish a gentle learning curve that leads to interaction beyond conscious control. Musical interfaces that cannot be physically apprehended tend to promote thinking processes rather than performative engagement. Certain operational complexity invites extended rehearsal that in turn makes the control seamless, once mastered.

2.2 Joystick following

An interesting interface for exploring sensor data processing is a simple joystick, as we experience it to be physically simple to describe in terms of tilt, while it is conveyed by two streams of data: the horizontal and the vertical position. These axes support the translation to polar coordinates that yields the radius i.e. the excursion from the centre and the gradient i.e. the angle of excursion, or tilt and rotation [1].

In this specific context of designing triggering models, a joystick that returns to the centre, due to springs attached, seems applicable for it's affordance to encode fast rhythmical gestures. Informal trials showed that striking the side of the joystick can be done with the accuracy and nuance normally associated with percussive instruments. Further, alternate striking back and forth can be performed at rather high pulsation rates. In order to implement sound generation for this type of interface a tracking model can be developed that can further facilitate articulation arising from skilled control that demands training.

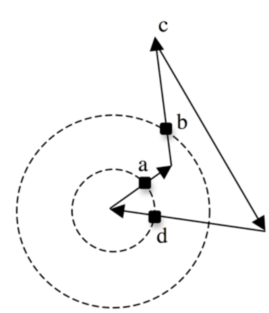

Figure 2. A complex joystick strike trajectory.

The core of this model is a state-following algorithm that proceeds to the consequent state in a circular configuration of 4 state transitions. First four letters of the alphabet in Figure 2 indicate these transitions. The two dashed circles represent the two threshold values for the excursion radius, while the arrows suggest the trajectory of the joystick movement during a complex strike.

This model operates as follows:

a) From the resting state, the transition a) will occur as soon as the radius supersedes a given, rather small threshold. This trigger is most useful for a fast-response attack synthesis. Measuring and inverting the time between the initial disturbance and the threshold crossing yields an estimation of the strike velocity.
b) A slightly larger threshold value is used in order to allow for analysis of further characteristics of the strike. Upon crossing this threshold the sustain stage of an instrument can be successfully triggered. Parameters for this trigger can be usefully derived from the timing and the angle of excursion.
c) The first occurrence of decreasing radius value can be considered a useful moment in time to initiate the release stage of an amplitude envelope. It can be characterised by the radius, angle and timing of this first local maximum. Another useful parameter to devise would be the difference to the angle of transition b), in order to estimate the amount of twist in the motion. Data streams representing the joystick position can, from this point on, be used to directly control the timbre of the sustained note.
d) Finally, the instrument can be brought to silence when the radius of joystick position decreases below the first threshold, closing the sequence of states allowing for a new initial trigger to occur.

This model of percussive interaction with a joystick seems to elevate the interface to new levels of complexity. Initial acquaintance with accurate rhythmical articulation will slowly reveal options residing in the model. The dynamics derived from transition timings and simple rotations of wrist already provide synthesis parameters that can yield a rather expressive instrument. Adding to this the twist before the release and many sustain articulations, it seems clear that the interaction model promotes virtuosity in performance and allows a learning curve that yields virtuosity that circumvents thought and becomes a physical skill that can be experientially related to the sound. Let this simple model exemplify a **state-following** strategy in designing complex interface-specific mappings for musical interaction.

2.3 Thresholding Foot Pressure

Interfaces measuring pressure exerted by dancers' feet on the floor could be used successfully for musical control as they capture the core of overall body movement in providing some data about motion, weight distribution and thereby the centre of gravity. In order to explore this, an interface that employs three pressure sensors was designed. The stepping surface was suspended such that all three point-sensors would provide a signal upon foot contact over a larger area. This provided an interface that was likely to benefit from arraying sensor data and designing a complex interaction model. In constructing a triggering model based on three pressure sensors the centre of gravity could be estimated by triangulation and similar techniques [4]. However, the accuracy of this particular sensor system did not allow for usable estimation of this position. Instead the position of contact was discretized to three possible areas by identifying the pressure sensors that exceeds the triggering threshold first. Once other sensors exceed the thresholds comparison of pressure magnitude is used to select the most depressed sensor. The model was further developed by introducing a 1st order, markov-style analysis [5], whereby the triggers are devised based on the previous and the current state of the sensors. This is shown in Table 2, where the first column lists the previous states, and the first row lists the current states of the system.

from\to	|000|	A>BC	B>AC	C>AB
|000|	0	1	2	3
A>BC	4	5	6	7
B>AC	8	9	10	11
C>AB	12	13	14	15

Table 2. 12 basic triggers devised from three pressure sensors.

The letter to the left of greater-than sign indicates the sensor with the highest reading, while |000| represents no-contact state. In this way 12 triggers are devised based on the trajectory of motion. The black shaded cells stand for triggers that can never occur, as the transition from a state to itself is impossible and is unaccounted for in the model. Further shading shows the difficulty of achieving specific triggers. It is fairly easy to initially trigger the interface in three different locations, which is represented by the second row. Also, there is no difficulty in removing all pressure from the interface after any of the three possible states. The remaining six triggers have proven to require some more skill as the centre of gravity needs to shift in a specific way. Overall, all 12 available triggers in this model were reliably used by a skilled dancer in informal trials.

Taking this model further, more than three states can be distinguished based on further comparison of the pressure measured at three points. In total, there are six possible configurations that can be assumed if the full order of pressure magnitudes is considered. These six, and the seventh state of no-contact, are represented in Figure 3 such that the first letter stands for the most pressured sensor and the third for the least. In this fashion, looking at the transition between two states a total of 42 different triggers can be devised using the model. Again, the diagonally positioned black cells represent impossible triggers as explained before.

from\to	|000|	ABC	ACB	BAC	BCA	CAB	CBA
|000|	0	1	2	3	4	5	6
ABC	7	8	9	10	11	12	13
ACB	14	15	16	17	18	19	20
BAC	21	22	23	24	25	26	27
BCA	28	29	30	31	32	33	34
CAB	35	36	37	38	39	40	41
CBA	42	43	44	45	46	47	48

Table 3. 42 triggers devised from three pressure sensors.

The assessment of difficulty for each trigger becomes more complex, but could be derived from the count of changing relationships between sensors that need to happen simultaneously. A single swap can be considered the easiest transition to obtain – cells representing these triggers have a white background. Slightly more difficult to achieve are triggers involving two place swaps in this symbolic representation using three letters – these are shaded light gray. Very difficult, if not impossible to achieve are the remaining six triggers. Before devising the difficulty of different triggers a probabilistic algorithm was employed to explore the likelihood of triggers occurring. Interestingly, a random walk (brown-noise) never obtained any of the difficult, dark-gray triggers while the proportion of the easy and medium triggers suggested the presented model to be appropriate.

The difficult triggers in this model could not be reliably used by a skilled performer. Complex two-feet gestures need to be devised to achieve this whereby the timing becomes paramount, as other transitions are very likely to intersect the trajectory attempted. This model is less usable for rhythmical triggering as the amount of triggers goes beyond what seems to fit into motor-memory of a performer. Instead, it is better suited to drive a more complex sound synthesis algorithms.

Important parameters that can describe the triggers are the velocity of movement and the total pressure measured. Using these two data streams with the two models described allow for creating an expressive instrument for dance controlled musical performance. Other quasi-continuous parameters can be devised to include the density of triggering, timing regularity, recurrence of patterns and many more. This case study, however, focuses on devising triggers and uses this final study to demonstrate a combination of dimensional recombination and sequence following strategies. Worthwhile noting is that these models can be more easily controlled from a seated position as the weight distribution of the foot on the floor can be more accurately controlled.

3. TRIGGER RELIABILITY

From the case studies presented we could conclude that the models designed relate to the way one can describe motion and as such similar models could be designed for other types of interfaces. This has been dubbed conceptual mapping in the literature [6], and covers strategies that employ a metaphorical approach that brings the design closer to the understanding of wider audience and user groups. These case studies were produced with simple sensors that may not be sufficiently accurate to achieve all possible triggers. Therefore, in order to increase the reliability of the model the first step is to choose and tune the sensor interface for maximum accuracy. However, there are further techniques that can help in building a reliable interface even when using simple sensors.

3.1 Timer Based Inhibition

Depending on the accuracy of the sensor interfaces the models described can suffer from multiple-triggers occurring in a very short amount of time. One of the solutions that can be employed is to implement a trigger inhibitor, or debouncing, which prevents further transitions from being derived for a certain period of time after the initial trigger has occurred.

3.2 Data Stream Lagging

The analysed streams of data can be processed to yield signals more suitable for given models. Basic lagging of the signals i.e. smoothing or low-pass filtering at sub-audio frequencies can prove very useful. Unfortunately this technique introduces some latency to the system and has to be used with caution. In certain instances when the direction of the crucially timed triggering is pre-defined, as in stepping onto a sensor, two-constant filters can be employed to lag the signal more in the less-crucial direction, akin to rise-decay filtering [7] applications.

4. NOTES ON VIRTUOSITY

Virtuosic performance is a phenomenon and audiences have an immediate, but potentially flawed sensation of the skills needed. Computer interfaces that are relatively easy-to-use often prompt performers to exaggerate the gestures to make the appearance of more effort. These might be considered extra-musical and unnecessary but are also considered sound accompanying, facilitating, ancilliary or performative [8]. This paper focuses away from this performative acting and strives to highlight ways of creating difficult-to-use musical instruments, Virtuosity, as it shows in performance in the realm of time-based art thrives on physical mastering of a difficult task. Mastering such a task in a musical fashion establishes a direct link between emotional intent and musical performance such that the complexity of the instrument and cognitive load are left out of the equation. Further aspects of how virtuosity comes to exist might help to establish how performance with digital instruments could induce a similar degree of awe.

4.1 Cultural Context

The technical command of the interfaces implied by virtuosity initially draws on difficulty of operation posed by the medium. However, this does not necessarily yield an exhaustive definition; if we consider a handclap, it is a simple acceleration and impact sequence that arguably all humans are capable of and unlikely would anyone consider it difficult. Nevertheless, people are likely to agree that there is some quite virtuousic hand clapping in the flamenco tradition. Therefore, it might be concluded that the frame of reference dictated by the tradition of performance plays a great role in our appreciation of extraordinary skill.

4.2 Beyond Simulation

Many digital gesture-following systems deal with analysis of motion that originate in performance of acoustic instruments. Inarguably that scientific endeavour brings valuable techniques and conclusions to the community. The question is, however, how can it stimulate the design of novel instruments and interaction strategies. Some experience with building augmented instruments has brought the author to the conclusion that the performers who play their instrument through an audio effect hardly develop novel musical expression compared to the performers who manage to circumvent their learned physical automatism and rediscover their own instrument. It seems thus that the existing motor-memory is detrimental in trying to explore the full potential of an instrument extension. Similarly, analysis of already acquired gestural skills seems inferior to building interfaces that necessitate learning of new skills, when the aim is to develop musical novelty.

4.3 Beyond Ease-of-use

Historically, the design of acoustic instruments might be considered to facilitate the control of a complex mechanism for producing audible vibration. This is clear if we consider, for example the keys of a clarinet or a cymbal stand. However, this tendency translates into a musical disaster in the digital age [9], exemplified by guitar-hero [10], which excludes both musical articulation and novelty. This statement is made to provoke an argument and it is not unjustified to consider this tendency a valid cultural development, let it thus remain an artistic choice in this context. Clearly, in trying to promote virtuosity by designing difficult-to-use interface data transformations, this paper argues that making instrument easy-to-use takes away from the virtuosity that can be accumulated. Nevertheless, it is not ruled out that guitar-hero platform cannot be taken to extreme modes of performance and artistic expression. Consider the turntable, a machine that made it rather easy to have an orchestra play in one's sitting room and how its use culminated in the art of scratching, which bears technical complexity that matches traditional acoustic instruments.

5. CONCLUSIONS

In designing sensor signal processing models two approaches seem to be opposed. An *analytic* approach would cover instances of which the starting point is a clear conception of what information needs to be extracted. The term *synthetic* approach may cover instances whereby the properties and affordances of sensor data inspire the model, or whereby the model is arrived at experimentally by using generic processing modules. Future research is aimed at producing, documenting and analysing a library of control signal processing modules that facilitate this *synthetic* approach. Extensive experimentation is facilitated by distinguishing events from streams and identifying myriad ways of transforming data between these domains.

But what does it actually take for an instrument to allow for virtuousic performance? It is the hype, is your author's humble suggestion. Most of our cultural development is just a product of our herd-instinct that makes us go in the same direction. Without the hype around violins, turntables or sequenced music, arguably none would enjoy the status that they currently have. It is thus not necessary to build difficult-to-use interfaces, a device with just a play/stop button was appropriated into a full-scale instrument (turntable). Indeed, in contemporary culture, the diluting effects of sensory saturation increasingly detaches the performance from complex musical control, just consider the expressiveness of a DJ dancing to a tune that he is 'playing'. The artistic experience therefore transcends the mediation that normally required isolated performance training, credit to the conformism of the masses and the inclusivity paradigm. The artistic experience is therefore likely to transgress mastery and come to reside with individual amateurish performance devoid of effort necessitated by traditional instruments. One can thus strive to be a hero expressing oneself with the aid of facilitating technologies or a fool deliberately putting complex obstacles in one's own way.

6. REFERENCES

[1] D. Wessel, and M. Wright, *"Problems and prospects for intimate musical control of computers,"* Computer music journal 26, 2002.

[2] J. C. Risset and M. Mathews *"Analysis of Musical Instrument Tones,"* Physics Today, 1969.

[3] Anonymised

[4] J. L. Poirot and G. Arbid. "Position Location: Triangulation Versus Circulation," IEEE Transactions on Aerospace and Electronic Systems, 1978.

[5] D.L. Isaacson and R.W. Maden. *Markov chains, theory and applications*. R.E. Krieger, 1986.

[6] C. Cadoz and M. Wanderley, *"Gesture–Music"*, M. Wanderley, M. Battier. Trends in Gestural Control of Music. IRCAM, 2000.

[7] U. Zolzer and X. Amatriain, *Digital Audio Effects*, Wiley, 2002.

[8] *R. I. Godøy and M. Leman, Musical Gestures Sound, Movement, and Meaning,* Routledge, 2010.

[9] T. Magnusson, *"Designing constraints: composing and performing with digital musical systems"*. Computer Music Journal, 2010.

[10] [6] http://www.activision.com/games/guitar-hero/guitar-hero-warriors-of-rock, accessed: 31/1/2014