Sound Spheres: the Heuristic Design and Evaluation of a Non-contact Finger-tracking Virtual Musical Instrument

Craig Hughes  
Michel Wermelinger  
Simon Holland

Computing Department and Centre for Research in Computing  
The Open University Walton Hall,  
Milton Keynes MK7 6AA, UK  
ch3375@student.open.ac.uk, m.a.wermelinger, s.holland @open.ac.uk

An important challenge in music interaction design is the creation of Virtual Musical instruments (VMIs) that offer expressive, playable, learnable mappings from gesture to sound, and which progressively reward practice by players. This design problem is particularly demanding in the case of non-contact musical instruments where there are limited physical reference points, and no haptic feedback. In this paper we gather existing criteria from the literature to assemble a simple set of light-weight design and evaluation heuristics for VMIs called articulacy. We present a design case study in which an expressive non-contact finger-tracking VMI, Sound Spheres, is designed and evaluated with the support of the articulacy heuristics. The case study explores the extent to which articulacy usefully informs the design of a non-contact VMI. We reflect on the possible usefulness of such heuristic approaches, and their limitations.

1. INTRODUCTION

With traditional musical instruments, there is a strong coupling between the playing gestures and the sound-producing mechanisms: the two exert strong constraints on each other. By contrast, in the case of Virtual Musical Instruments (VMIs) [4,6] interaction gestures and sound design are, in principle, independent. Consequently, the design of VMIs requires explicit attention to the mapping from gesture to sound.

Despite the freedom thus afforded, a review of sources such as Mulder [3] and the Taxonomy for real-time Interfaces for Electronic Music performance (TIEM) [5] suggests that the majority of VMI controllers nevertheless rely on physical interaction between player and instrument. This is far from surprising. Research in Physicality in Human Computer Interaction [10] and embodiment in Music Interaction Design [9] suggest numerous affordances offered by physical contact.

However, some VMIs are controlled without physical contact interaction [3,5] and instead rely on free-air gestures. Non-contact VMIs raise interesting challenges for music interaction designers.

Interaction designs for VMIs are often arrived at intuitively. By contrast, some designers find design and evaluation heuristics [11] useful. This paper reports on a design case study in which an expressive non-contact finger tracking VMI is designed and evaluated using a candidate set of design heuristics and evaluation heuristics for VMIs. These heuristics, which we have labeled articulacy, (defined in section 3.1 below) are derived from design considerations from the literature [1,2,4,6]. The present case study considers how design and evaluation heuristics such as articulacy can inform a non-contact VMI design. We offer a preliminary reflection on the usefulness of such heuristics for this purpose, and their limitations.

2. DESIGN CRITERIA

VMIs that successfully appeal to performers involve rich and subtle constraints on the connections between gesture and sound. However it is hard to characterize this explicitly. The HCI literature suggests many candidate design considerations, some relatively simple, such as clarity of feedback [11], and others more complex, such as appropriate exploitation of physicality [10] and systematic consideration of issues of embodiment [9]. For the present purposes, a set of simple, light-weight considerations were chosen.

The Thummer Mapping Project [4] identified four common physical instrument variables (pressure, speed, angle and position) that control instrument dynamics, pitch, vibrato and articulation. In a later study Paine [6] re-iterated these control parameters as important factors for the design of new musical interfaces. Jordà [1,2] described other factors considered important to the consideration of a good musical instrument, suggesting playability, progression (learning curve), control and predictability. He also suggested that the balance between challenge, frustration and boredom must be met. Ferguson and Wanderley [8] highlighted reproducibility as one more important factor for digital musical instruments, suggesting that musical instruments that allow a performer to be expressive must also permit a performer to imagine a musical idea and be able to reproduce it.

We have borrowed and adapted these various considerations to provide a simple lightweight set of heuristics for the design and formative evaluation of non-contact VMIs.

2.1 Articulacy heuristics

We will consider the articulacy of a non-contact VMI to refer to (a) the degree to which pressure, speed, angle and position can be used to control the instrument and
(b) the degree to which the design achieves playability, progression, control, predictability, reproducibility, and balance between challenge, frustration and boredom.

This set of considerations can be applied straightforwardly to VMI design simply by using them as a checklist of desirable properties. Similarly, they can be applied to formative VMI evaluation by considering, or measuring (see section 6), the extent to which they are achieved in a given design. Despite the extreme simplicity of this method, similarly simple approaches have been found useful elsewhere; for example, our approach broadly echoes such approaches as Molich and Nielson [12].

3. OVERVIEW OF SOUND SPHERES

The Sound Spheres VMI is controlled solely by finger movement in the air. Unlike some finger tracking applications, complex finger gestures are avoided and only the finger tips are used. Highly reflective tape placed on the fingertips reflects infrared light to the Wiimote’s infrared camera (figure 10). The Wiimote then passes data concerning the positioning of the fingertips to the Sound Spheres VMI software.

The position of the finger tips is represented on the user interface (figure 1) as small spheres (tracking spheres). Only four fingertips can be simultaneously tracked with the Wiimote’s infrared camera and hence this poses a limitation of up to a maximum of four tracking spheres. The movement of the tracking spheres is used to trigger sounds through collision with a set of fixed larger spheres (the sound spheres), which are organized in two rows, each comprising the 12 notes of an octave (figure 1). The two rows correspond to two different octaves, one octave apart. To differentiate the natural notes from sharp notes, sound spheres of different sizes are used.

4. DESIGN OF SOUND SPHERES

To support the design of the Sound Spheres VMI we used the articulacy design heuristics outlined above to guide a rapid prototyping approach.

Given the starting point – fingers in free air directing the collision of spheres to produce sounds – there are, broadly speaking, three principal categories of design decision to be made, summarized in table 1.

The first is the design of specific gestures, or aspects of gesture, for each of the four instrument control parameters identified by articulacy, i.e. position, angle speed and pressure. In practical terms, this decision particularly concerns how the values of the various control parameters are to be derived from the finger tracking data. The second category of design decision is to map each control parameter to an appropriate sound shaping operation. The third is to design visual feedback as needed.

In the remainder of this section, we will outline the principal design decisions associated with each of the four control parameters in turn (sections 5.1 - 5.4) and then consider visual feedback for the VMI as a whole (section 5.5).

<table>
<thead>
<tr>
<th>Instrument control parameters</th>
<th>Effect on sound</th>
<th>Visual feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Stereo Panning</td>
<td>Flying Sparks</td>
</tr>
<tr>
<td>Position of a tracking sphere at point of collision (figure 3).</td>
<td>Sound is increasingly panned to left or right speaker dependent on position of collision.</td>
<td>Direction of sparks depends on the position of tracking sphere collision (fig. 6).</td>
</tr>
<tr>
<td>Speed</td>
<td>Volume</td>
<td>Spin</td>
</tr>
<tr>
<td>Speed of tracking sphere’s movement at point of collision (figure 4).</td>
<td>A greater speed results in a higher volume.</td>
<td>The greater the speed of the tracking sphere the faster the visible spin on collision.</td>
</tr>
<tr>
<td>Pressure</td>
<td>Parametric EQ</td>
<td>Size</td>
</tr>
<tr>
<td>Based on momentum of tracking sphere at point of collision. Tracking sphere size is changed to increase or decrease momentum.</td>
<td>A greater pressure results in a tone where the higher frequencies are boosted.</td>
<td>The greater the pressure the larger the tracking sphere.</td>
</tr>
</tbody>
</table>
Challenge, an analogy with piano fingering suggests that progression and chaos are broadly welcome design trade-offs. For example, piano or xylophone manufacturers have an interest in prominent surface resonance that is enhanced by the resonance of adjacent sound spheres. However, there is a need to strike sound spheres at a point that is defined by the sound sphere’s central line to achieve the desired result. Thus, when a tracking sphere collides with a sound sphere the point of collision is adjusted dependent on the acute angle of collision and the distance from the point of collision. The sound generated at the point of collision is adjusted by the time difference between the start and collision positions as illustrated in figure 4. The speed is used to adjust the sound generated at the point of collision simply by adjusting the volume, with a greater speed resulting in a higher volume.

Prompting design decisions associated with instrument control parameters is helpful, but the articulacy heuristics also prompt a consideration of the degree to which any design decisions impact on playability, progression, control, predictability, reproducing abilities, and balance between challenge, frustration, and boredom.

In the case of speed, articulacy’s playability design heuristic prompts the question of how it might be possible for a player to execute low speed gestures when there is a need to strike sound spheres rapidly in succession. However, reflection reminds us that an analogous situation exists in many traditional instruments without playability being impaired, for example piano or xylophone. Of course, playability depends on skill, but the present design appears to offer a broadly welcome design trade-off between playability, progression and challenge.

Continuing the prompted reflection on playability and challenge, an analogy with piano fingering suggests that the Sound Spheres player has a choice of playing a forthcoming note with any of the four tracking spheres and hence finger distance could be minimized with practice.

Table 1. Outline of principal design decisions

<table>
<thead>
<tr>
<th>Angle</th>
<th>Chorus</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle generated by a tracking sphere’s start position and collision point (figure 5).</td>
<td>An acute angle results in a chorus effect with a greater degree of modulation than a less acute angle.</td>
<td></td>
</tr>
</tbody>
</table>

4.1 Key design decisions for Position

When a tracking sphere collides with a sound sphere the value of the position control parameter is taken to be the horizontal distance from the point of collision and the central line of the sound sphere. This position is used to modify the sound generated at the point of collision by stereo panning to the left or right according to the distance from the sound sphere’s central line (see figure 3 and table 1).

Figure 3. Position articulation

4.2 Key design decisions for Speed

When a tracking sphere collides with a sound sphere the average speed of the tracking sphere is taken to be the distance between the start and collision positions divided by the time difference between the start and collision positions, as illustrated in figure 4. The speed is used to adjust the sound generated at the point of collision simply by adjusting the volume, with a greater speed resulting in a higher volume.

Figure 4. Speed articulation

Finally, a small movement of the fingers can affect a big movement in the tracking spheres (sensitivity) allowing individual adjustment of the “action” of sound spheres to assist playability.

4.3 Key design decisions for Angle

Compared with the design decisions associated with position and speed, the design decisions for angle are necessarily a little more oblique. The key facilitating step turned out to be to consider the starting position for a finger trajectory, as well as the collision point.

Thus, when a tracking sphere collides with a sound sphere the angle is taken to be the acute angle between three points, as illustrated in figure 5: point 1 is the center of the tracking sphere at the start of its movement towards the sound sphere, point 2 is the center of the tracking sphere at the point of collision with the sound sphere, and point 3 is any point horizontally displaced from the point of collision. The sound generated at the point of collision is adjusted dependent on the acute angle between these points. A more acute angle results in a chorus effect with a greater degree of modulation than a less acute angle.

Thus, the collision of a tracking sphere with a sound sphere at an identical position can sound different depending on the starting position of the tracking sphere. This enables musical expression by swiping fingers in different ways. The articulacy heuristics again direct us to consider the degree to which this design decision may impact on such factors as playability, progression, control, predictability, reproducing abilities, and the balance between challenge, frustration, and boredom.

This leads to a reflection on analogous situations, such as when a drummer strikes a cymbal. A change in the angle at which a drummer strikes a cymbal will produce a different sound. Sometimes a player will use a shallow angle and appear to brush the drumstick over the surface of the cymbal, and sometimes a more direct hit is executed, with widely different sounds being generated. Returning to the design decision in Sound Spheres, little
can be concluded about playability, but these considerations do suggest challenge and possible progression.

The method outlined above for determining the angle of a collision assumes that the starting position of each finger-driven tracking-sphere trajectory is well defined. In practice, the transition of a tracking sphere from playing one note to the next will frequently involve continuous motion, and hence the point at which a movement corresponds to the start of playing a new note can be difficult to ascertain. The engineering decision as to how the starting point is to be identified will have implications for articulacy factors such as playability, so reflection on playability is prompted. When playing a traditional percussive instrument such as a xylophone or steel drums, or even a stringed instrument like the piano, the movement of the striking object (be it a mallet, stick or fingers) from one note to the next is rarely linear. A player generally lifts the object from one striking position before they start the movement to make another strike. With this in mind, the decision was taken for the tracking sphere’s starting position to be determined by the point at which the movement changes from a positive direction in the y-plane to a negative one, i.e. the point at which a downward movement begins, following an upward movement.

4.4 Other key design decisions: Pressure

In a non-contact environment, finding an appropriate gesture, or aspect of gesture, to map onto a pressure control parameter presents a design challenge.

To help guide design, pressure was deemed to be closely related conceptually to momentum. If we assume the virtual mass of tracking spheres to be proportional to their size, we can conclude that a larger tracking sphere would exert a greater pressure on a sound sphere than a smaller tracking sphere travelling at the same velocity. In other words, by varying the size of a tracking sphere we can vary the pressure being applied to a sound sphere during collision.

To implement the ability to dynamically and rapidly change the size of the tracking spheres, the user interface displays a visual component called a pressure control (Figure 1). A pressure control has been placed on either side of the user interface so that it can be quickly accessed by tracking spheres controlled by either the player’s right or left hand. The pressure control has two circular surfaces, one containing an upwards facing arrow representing increasing pressure and one a downward facing arrow representing decreasing pressure. This control will gradually increase or decrease the size (and hence the implied pressure) of all tracking spheres when the center point of one of the tracking spheres is positioned over one of the pressure control’s surfaces.

Reflecting once more on playability, progression and challenge, it is clear that by using one hand to vary pressure while the other hand triggers sounds, it should be possible to change pressure relatively rapidly.

Pressure is used to modify the sound generated at the point of collision in the following way: a greater pressure results in a tone where the higher frequencies are boosted using parametric EQ.

4.5 Visual feedback

The articulacy heuristics encourage the use of visual feedback to assist with the communication of position angle and speed, moderated by considerations such as playability, progression, control, predictability, reproducibility, challenge, frustration and boredom. The Sound Spheres VMI provides visual feedback to the player when the tracking spheres collide with sound spheres in several ways, as follows.

Firstly, graphics are displayed at the point of each collision (figure 6). A graphics particle engine was implemented to display a set of flying sparks at the point of collision. The direction and dispersal of the sparks is dependent on the position of the collision in the sense defined in section 5.1. This is illustrated in Figure 6.

Figure 6. Sphere Collision Sparks

Secondly, when a tracking sphere collides with a sound sphere, the sound sphere vibrates as if it were on a spring. The vibration diminishes over time and then stops. The direction of the vibration is always up and down. Consideration was given as to whether the direction of vibration should also be dependent on angle, however reflection on articulacy issues prompted this idea to be dropped. As the sound spheres are placed close together, any sideways vibration could result in their collision, with likely negative consequences for playability, control and frustration. Hence, the vibration of the tracking spheres is not related to any specific control parameter and denotes sphere collision only.

Craig Hughes, Michel Wermelinger and Simon Holland

Sound Spheres: the Heuristic Design and Evaluation of a Non-contact Finger-tracking Virtual Musical Instrument
was asked to complete a questionnaire with 49 questions. The initial 5 questions served to identify the participant and their ability to play and read music. Two questions asked the participant to rank the control parameters in terms of ease of use and importance to musical outcomes. The last 3 questions asked for general comments about what participants liked most and least about Sound Spheres. The remaining 39 questions covered the various design factors (playability, progression, control, predictability, reproducing and, balance between challenge, frustration and boredom), asking participants to respond using a 5-point Likert rating scale (strongly disagree, disagree, neither agree or disagree, agree, and strongly agree) thus providing quantitative data to which statistical analysis could be applied.

Spearman’s rank correlation method was used to determine the relationship between 57 pairs of questionnaire responses, e.g. if the ease of use of the speed control parameter correlated with the preference for its applied visual feedback. Furthermore, due to the small sample size, the non-parametric Mann-Whitney U Test was systematically applied to each of the 41 questions to test the hypotheses that questions may be answered differently between musicians and non-musicians, and between those who did and did not participate in the prototype reviews.

6. RESULTS

Statistical analysis of the questionnaire responses showed strongly positive feedback to many factors relating to the Sound Spheres VMI. For example, 87.5% of participants thought that the Sound Spheres VMI facilitated the creation of music well and that their playing improved over time. 75% of participants thought that it was easy to move the tracking spheres using the finger tracking method. Responses to questions about factors such as general playability, the progression of the musician’s ability, control, and balance between challenge, frustration and boredom suggested that the Sound Spheres VMI was generally judged positively in these respects. Responses to questions on the factors of predictability and reproducibility generally showed negative judgments in these areas. However, observation and the results of the Mann-Whitney U Tests suggested less negative judgments where more playing time (i.e. practice) was given to the participants.
The control parameters of pressure, speed, angle and position were ranked from 1 to 4 based upon their ease of control (1 being the easiest and 4 being the hardest) and also for their importance to musical outcomes, i.e. which control could be used best for affecting the musical outcome (1 being the most important and 4 being the least). A scoring system was applied to the rankings received by each of the participants (4 points were given to a rank of 1, 3 to a rank of 2, etc.) and a ranked scoring was calculated for each control parameter. The percentage of the sum of all the control parameters scorings was calculated for each. These percentages are shown in Figure 8. In general the control of pressure, speed, and position was considered easy, and the sounds generated for each of these controls were considered apparent, consistent and appropriate. Angle was the control parameter that received the most negative feedback in terms of its ease of control and associated audio result.

There appear to be several reasons for this, which we will briefly review. Firstly, the positioning of the sharp note sound spheres (which were placed lower than the natural notes) made them difficult to hit at an angle. Secondly, participants found that they often played more than one intended note when using the angle control due to the close proximity of sound spheres. Thirdly, visual feedback was not implemented for the angle control parameter. This suggests the combination of both audio and visual feedback (synchresis) may play an important role in non-contact VMIs.

Only 8 of the 57 Spearman’s rank correlation results showed statistical significance and through further analysis 5 of these results were considered unreliable. For example, one negative correlation coefficient value suggests that the Sound Spheres VMI facilitates the creation of music better as the control of the tracking spheres gets harder. This is the reverse of what would be expected, especially considering that 87.5% of participants thought that the Sound Spheres VMI facilitated the creation of music well and 75% thought that the movement of the tracking spheres was easy. However a strong correlation exists between the improvement of ability to play the Sound Spheres VMI over time and the ability to distinguish the application of more than one control parameter at a time. This suggests that progression of ability or skill in playing the Sound Spheres VMI can be achieved. Correlation also suggests that accuracy in positioning the tracking spheres increases as the consistency in control of tracking sphere movement increases.

The Mann-Whitney U Test results indicated that there was no significant difference between musicians and non-musicians in the way questions were answered. However, there were five questions that identified significant (i.e. p < 0.05) differences between the responses of those who participated in the prototype review sessions and first time users of the Sound Spheres VMI. These results indicate that participants of the prototype review sessions were more able to consistently control the movement and position of the tracking spheres. They also used the control parameters to add expression during play more than first time participants. Participants of the prototype review sessions more strongly agreed with the change in sound being apparent and consistent when using the pressure control.

7. IMPLEMENTATION

The Sound Spheres VMI comprises the following components:
- The Sound Spheres software.
- Laptop computer and 24-bit sound card, external speakers and wide-screen monitor.
- Bluetooth adapter and supporting driver.
- Wiimote controller.
- Infrared LED array with cover.
- Four reflective markers.

The components are setup on a two-tiered desk with the top tier used as a surface on which to stand the speakers and computer monitor and the lower tier used as a surface for placement of the Wiimote and LED arrays. Separate tiers enable the Wiimote and LED arrays to be positioned horizontally central to the monitor and speakers without obstructing the player’s view of the monitor. The Wiimote and LED array can be adjusted up or down to suit desired playing positions. An adjustable chair also allows players to raise or lower their playing position. The reference speakers are positioned either side of the monitor so that stereo effects are maximized. The system’s setup can be seen in Figures 9 and 10.

8. LIMITATIONS

Articulacy is simply one example of a heuristic approach. The evaluation was carried out using subjective measures for the various articulacy characteristics.
Articulacy might be argued to have a potentially 'reactionary' influence on VMIs, since it focuses attention on relatively well-explored conceptual metaphors [9] such as position and speed. However, where this might be judged inappropriate, the present study offers a good starting point for constructing alternative sets of heuristics with contrasting characteristics.

9. RELATED WORK

Wanderley and Orio [13] carried out a systematic review of existing mainstream HCI techniques for evaluating input devices and considered how these might be applied to music interaction. They also explored the notion of benchmarks (i.e. common musical tasks) that might potentially form part of a task-centric evaluation methodology. Kiefer et al. [14] reviewed some newer developments, and reported on a case study which stressed the value of interview data for identifying unexpected usability issues. Seago [15] critiqued existing user interfaces for timbre design from an HCI point of view. Wilkie et al. [9] introduced a novel approach to evaluating user interfaces for music, based on embodied cognition, image schemas and conceptual metaphor. Holland et al. [16] outlined how this approach might be applied to whole body and other non-contact interfaces for music. Modhrain [17] offers a valuable reflection on the role of evaluation in Digital instrument design.

10. CONCLUSIONS

Design for non-contact VMIs is challenging. We have presented a case study demonstrating how a heuristic approach was used to design and formatively evaluate a novel non-contact VMI, Sound Spheres.

The present work provides a valuable cue for reflection in the context of the present workshop. The first NIME workshop (New Interfaces for Musical Expression) took place at CHI 2001 in Seattle, at a time when Second Paradigm of HCI (with its emphasis on tasks and human information processing) was ascendant. Apparently, the organizers of NIME found this approach so alien that they decamped and re-founded NIME as a separate conference. And yet, in 2001, Paul Dourish's 'Where the Action Is' crystallized many of the insights leading to HCI's Third Paradigm, with its focus on the embodied, the ludic, the experiential and the situated. This family of approaches seems especially well suited to the needs of Music Interaction Design. However, music demands that open-mindedness of approach will always remain essential.

11. ACKNOWLEDGEMENTS

This paper is a shortened and adapted version of [18].

12. REFERENCES


