ORIGINAL ARTICLE

Analyzing journeys in sound: usability of graphical interpolators for sound design

Darrell Gibson[1] · Richard Polfreman[2]

Abstract

Graphical interpolation systems provide a simple mechanism for the control of sound synthesis by providing a level of abstraction above the engine parameters, allowing users to explore different sounds without awareness of the underlying details. Typically, a graphical interpolator presents the user with a two-dimensional pane where a number of synthesizer presets, each representing a collection of synthesis parameter values, can be located. Moving an interpolation cursor within the pane results in the calculation of new parameter values, based on its position, the relative locations of the presets, and the mathematical interpolation function, thus generating new sounds. These systems supply users with two sensory modalities in the form of sonic output and visual feedback from the interface. A number of graphical interpolator systems have been developed over the years, with a variety of user-interface designs, but few have been subject to formal user evaluation. Our testing studied both user interaction with, and the perceived usability of, graphical interpolation systems by comparing alternative visualizations in order to establish whether the visual feedback provided by the interface aids the locating of sounds within the space. The outcomes of our study may help to better understand design considerations for graphical interpolators and inform future designs.

Keywords Sound · Synthesizer · Interpolation · Visualization · Interface · Sound design · Usability

1 Introduction

The fundamental problem when designing sounds with a synthesizer is how to configure the often large number of complex synthesizer parameters to create a certain audio output, i.e., how to translate sonic intent to parameter values. Although having a direct access to every parameter (*one-to-one* mapping) gives fine control over the sound, it can also result in a very complex sound design process that often relies heavily on experience; hence the widespread use of preset sounds and samples. Alternatively, it is possible to map a smaller number of control parameters to a larger number of synthesizer parameters (*few-to-many* mapping) in order to simplify the process [1]. Particular states of synthesis parameters (presets) can be associated with different control values, and then, as these control values are changed, new synthesizer parameter values are generated by interpolating between the values of the presets. This provides a mechanism for exploring a defined sound space, constrained by the known sounds, the visual model and the changes of the control parameters. The importance of mapping strategies has been previously recognized in the design of electronic musical instruments [1, 2], and it would seem likely that it is equally important to a sound design context. However, no formal studies appear to have been undertaken.

A number of such interpolation systems have been proposed over the years, and these can be categorized based on whether the control mechanism is via a graphical interface or in some other form. Those that are of interest here are those that offer a graphical representation that allows the control of a visual model.

✉ Darrell Gibson
dgibson@bournemouth.ac.uk

Richard Polfreman
r.polfreman@soton.ac.uk

[1] Faculty of Science & Technology, Bournemouth University, Dorset, UK

[2] Faculty of Arts and Humanities, University of Southampton, Hampshire, UK
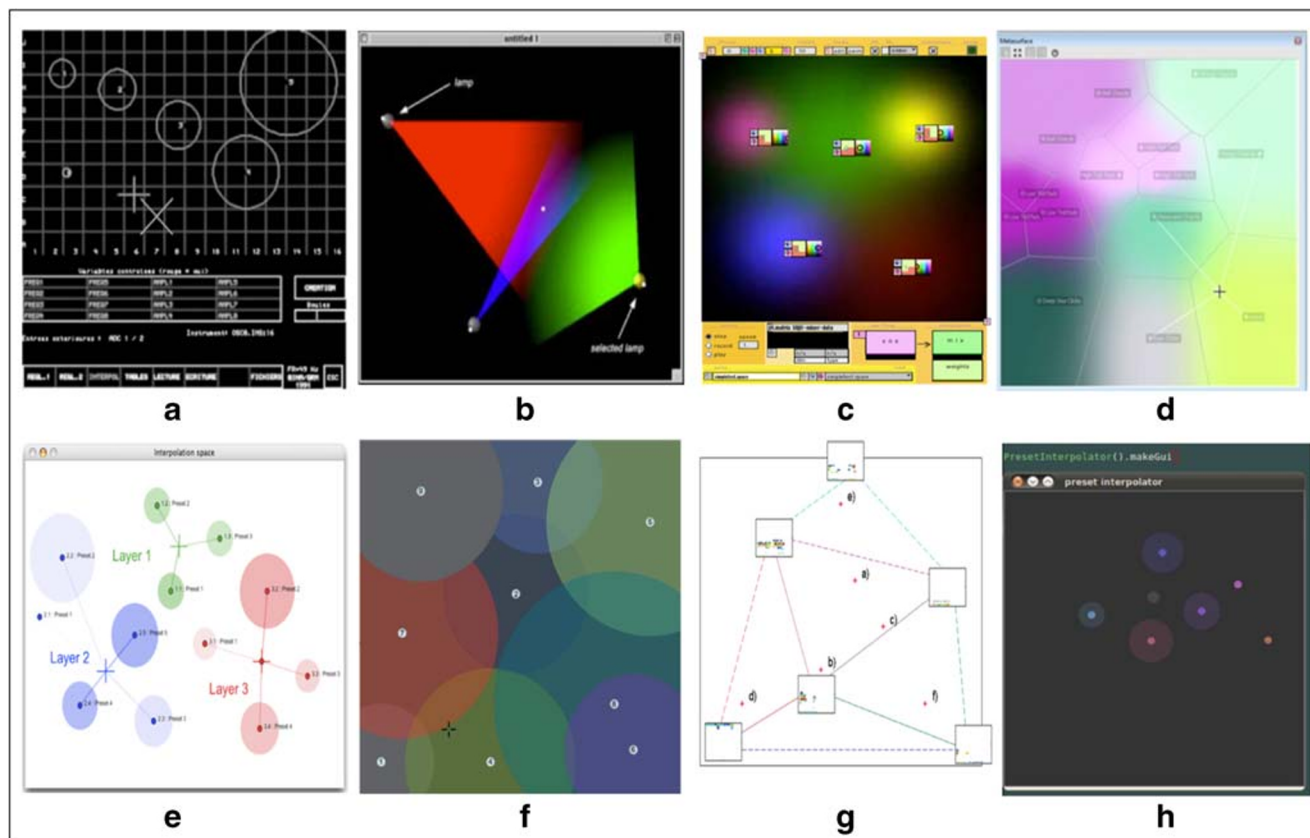
1.1 Graphical interpolation mapping

Graphical interpolation systems provide a display within which visual markers that represent presets can be positioned. Interpolation can then be used to generate new parameter values in between the specified locations by moving a cursor around the space. This allows the user to explore the interpolation space defined by the presets and discover new sonic outputs that are a function of the presets, their location within the interpolation space and the relative position of the interpolation point [3]. In addition, the precise visual model that the space employs will also affect the sounds that can be generated.

A number of these systems have been developed by researchers using a variety of distinct visual models for the interpolation [3–10], presenting the user with varying levels of visual feedback.

The systems shown in Fig. 1 highlight the range of different visual models that have been used for interpolation and corresponding visualizations. A more detailed review of these has been undertaken [11]; however, from this, it is difficult to gauge if the visual information actually aids the user in the identification of desirable sounds,

given that the goal is to obtain a certain sonic output, or whether the visual elements actually distract from this intention. Moreover, if the visualizations do aid the process, how (much) do they help and which visual cues will best serve the user when using the interface for sound design tasks?

Although, there have been a couple of comparative studies relating to interpolated parameter mappings in musical instrument design, none of these have focused on the visual aspects [12, 13]. Another area where spatial interpolation has been used is within Geographical Information Systems (GIS) to estimate new values based on the geographical distributions of known values. A review of these techniques is available [14], but the visualizations tend to be geographical maps and the interpolation is typically performed on either a single or few parameters, rather than the relatively large number used in synthesis. In an even wider context, general guidelines exist for the design of multimodal outputs [15], in our case audio and visual, but these are fairly broad and not specific to this class of system. Our work aims to provide the missing link and specifically focus on the interpolator usability in sound design, based on the visual display and its combined effect with the sonic output.
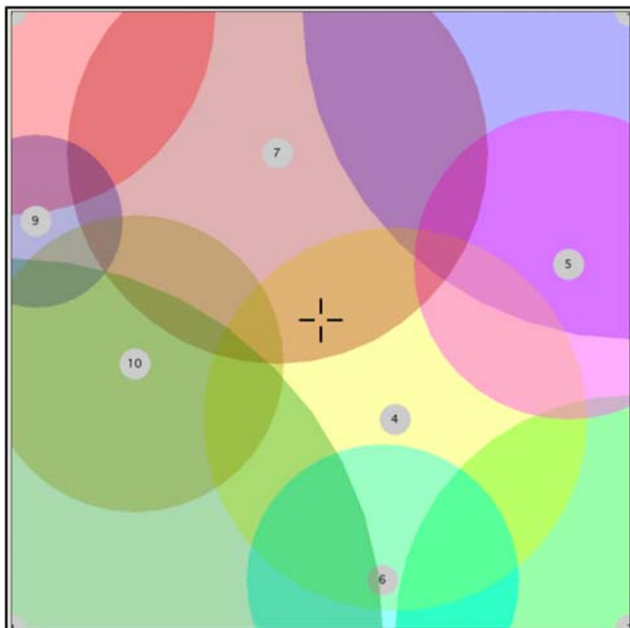

Fig. 1 Graphical interpolator models. **a** SYTER [4], **b** interpolator [5], **c** Gaussian kernels [3], **d** metasurface [6], **e** INT.LIB [7], **f** nodes [8], **g** Delaunay triangulation with spikes [9], and (f) intersecting N-spheres [10]

1.2 A typical interpolator: nodes

A readily available visual model for a graphic interpolator is realized in the visual programming environment Max, a software package designed for developing interactive audio and/ or visual systems. Andrew Benson, a visual artist, created the *nodes* object in 2009, and it proved so popular that it has been included in subsequent Max distributions [8]. This visual model uses a location/distance-based function, where each preset is represented as a circular node within the interpolation space. The size of each node can be set and defines the node's extent and so influence within the interpolation space. Figure 2 shows an example of an interpolation space created with the nodes representation.

With this model, the weightings are calculated when the cursor coincides with a node's location, and the weighting is calculated as the normalized distance to the node's center (0.0–1.0). That is, when the cursor is at the node center a weighting of one is generated and the value linearly reduces to zero at the node boundary. The interpolation is then performed where multiple nodes intersect. For example, in Fig. 2, the cursor position shown corresponds to the overlap of nodes 4 and 7 with the shown relative weights. Therefore, the interpolation is performed based on the normalized weighted distance of any containing nodes, and the size and relative position of each node determines its influence. The node weightings are updated in real time as the interpolation point is moved or if the nodes are resized or repositioned within the space. When the interpolation point is on an area only occupied by a single node, then the preset that corresponds to that node will be recalled.
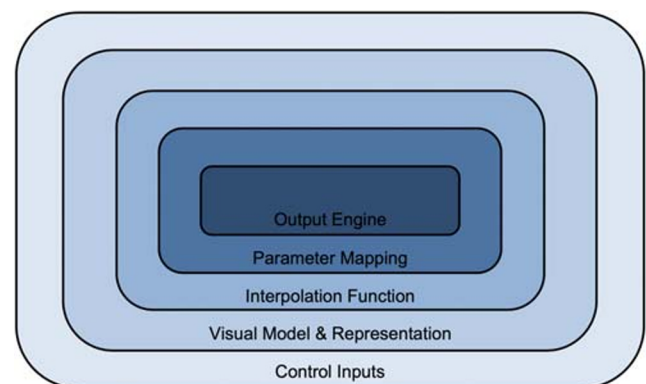
It should be stated that although we are specifically interested in the control of sound synthesis, the nodes object can be used for interpolation of any desired parameters, e.g., for computer graphics, animation, color, or sound effect processes.

2 Generic interpolation framework

In order to evaluate different visualizations, an interpolation system was defined that allowed the modification of the visual representation for the interpolation model, while leaving all other aspects the same. However, interpolation systems actually consist of five separate elements, each of which can potentially change the functionality and usability of the system.
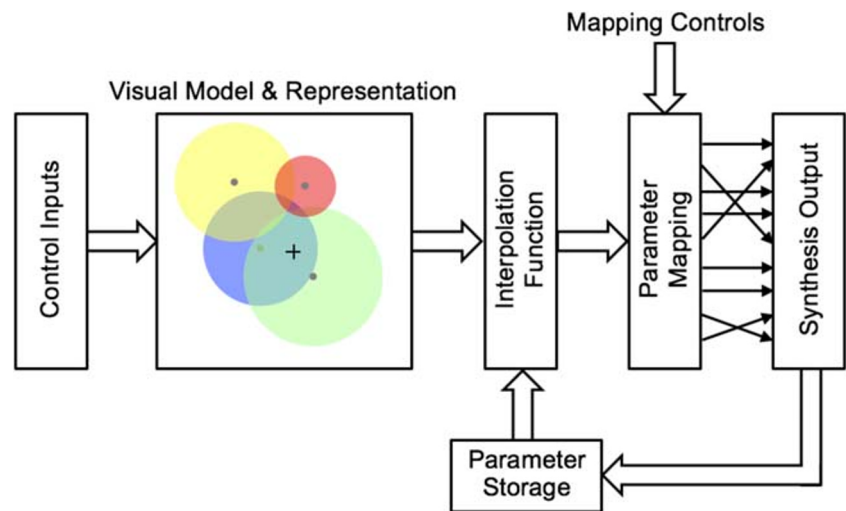
1. *Control* The input controls of the interpolation model
2. *Visualization* The interpolation model and how it is represented
3. *Interpolation* The actual interpolation calculation
4. *Mappings* Between the interpolation and the synthesis parameters
5. *Engine* Type/architecture/implementation of the output engine

While this paper focuses on the visualization, the other components needed to be implemented so that the different visualizations could be investigated. To understand the hierarchical relationship of each and the impact on the output, a conceptual framework has been previously presented [11]. This has subsequently been refined to become a generic framework for all interpolation systems, regardless of the actual application area, e.g., computer graphics, animation, color, or sound effects processes. As shown in Fig. 3, this begins with the control inputs at the top level and works down to the output engine at the bottom. Design changes at any stage in the pipeline will impact on all subsequent stages and will

Fig. 2 Interpolation space created with Max nodes object with the weightings distances of 0.355 for node 4 and 0.180 for node 7 for shown cursor position

Fig.3 Generic interpolation framework showing the hierarchical relationships for each stage and interdependencies

effectively change the interpolators output. In our case, this
will change the potential sonic pallet that the interpolator
provides.

Although the final output is at the bottom level, it is worth
noting that the displayed representation gives the user visual
feedback on the current configuration of the interpolation system
and potentially the output sound. Equally the user may be given
inputs that allow the mappings to be modified. However, the
framework shows the interdependencies of the different elements
of an interpolation system and the relationships between them.

2.1 Framework implementation

An implementation of a graphical interpolator was built that
used the framework to structure the different levels (control,
visual model, interpolation, mappings, and synthesis) into
separate modules. This allowed a single module in the frame-
work to be modified independently of the others, and the
outcomes compared. This facilitated comparative user testing,
where the results could be measured, analyzed, and evaluated
to gain a better understanding of the important factors when
designing new graphical interpolation systems. The Max vi-
sual programming environment was used to construct the ar-
chitecture shown in Fig. 4.

As the Max *nodes* object has been specifically designed as
a graphical interpolator, it already contains specific function-
ality for the *visual model* and the *control input* elements of the
framework. The *control inputs* realized in the *nodes* object are
standard computer-based spatial controls, e.g., moving the
cursor with a mouse. The interpolation point on the *nodes*
object can be moved within the space, and a normalized
weighting is generated for each node that the cursor intersects.
All the weightings are used as the input to the *interpolation
function*. The *interpolation function* module receives data
from storage that holds all the parameter values for each pre-
set. All the weightings are summed and the individual

weightings are normalized relative to this total to give each
as a percentage, as shown in Fig. 5. These are then used to
determine the relative values of each preset parameter before
they are sent to the synthesis engine. By default, the calcula-
tion of the parameters is performed using a linear interpolation
function, but it is possible to change the mode so that alterna-
tive functions can be used.

The *synthesis engine* has been constructed to be separate
from the interpolation platform by using software plug-ins.
This allows different (including commercially available) syn-
thesis engines to be used with the interpolator. When a new
synthesizer is loaded, it is interrogated to determine all the
parameter values for the selected presets. Each preset is
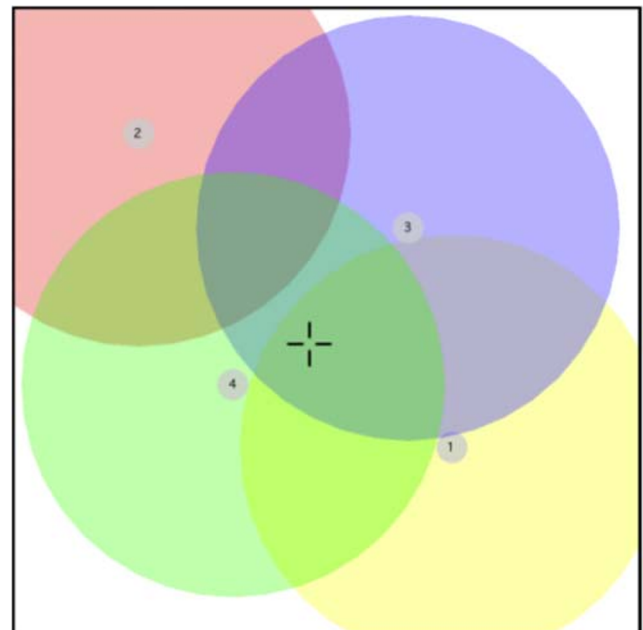
Fig. 5 Nodes outputs weighted distances resulting in $1 = 0.1696$ (16%),
$2 = 0.0000$ (0%), $3 = 0.3180$ (30%), and $4 = 0.5724$ (52%) for shown
cursor position from the interpolation function

associated to a node in the interpolation space and the preset's parameter values are sent to the *interpolation function* storage.

By default, all of the parameters for the presets are associated to the corresponding node and so every aspect of the sound's synthesis is controllable. However, the *parameter mappings* between the *interpolation function* and *synthesis engine* can be changed by the user through the *mapping controls*, permitting different mappings to be selected. The controls currently allow subsets of the available parameters to be selected, but these could be modified to allow more complex mapping strategies.

2.2 Nodes reimplementation

As the source code for the *nodes* object is not available, it needed to be reimplemented so that the visualization of the model could be customized for comparative testing. The *nodes* object was replaced with an interactive user interface that was built using OpenGL for implementing the interpolation model's visual representation and JavaScript to define the control mechanism and to generate the preset weightings. This separation allowed the testing of different visualizations with the same nodes interpolation model to be carried out while also facilitating implementation of alternative interpolator models in further studies.

The reimplementation of the nodes model was functionally tested by undertaking back-to-back tests between it and the original nodes object, ensuring that both implementations gave the same results.

3 Experiment design

Using the framework structure detailed in the previous section, an experiment was designed to evaluate user interactions with the graphical interpolation systems, where different levels of visual feedback could be supplied [16]. The aim was to evaluate the user interactions to determine the impact of the visualization on systems usability. To assess the usability of the interface five metrics were identified for investigation:

1. *Time* Do different visualizations encourage users to spend more time exploring the interpolation space?
2. *Speed* Do the visualizations presented support users to move faster in the interpolation space?
3. *Distance* Do the presented visualizations facilitate the travel of longer distances of the interpolation space?
4. *Accuracy* Does the visualization allow users to be more accurate when locating sounds in the space?
5. *Satisfaction* Do the visuals presented affect the perceived operation of the system?

The ISO 9241-11 (1998) standard defines the usability of systems in terms of *efficiency*, *effectiveness*, and *satisfaction* [17]. The metrics time and speed were chosen to measure efficiency, distance and speed for effectiveness, and then satisfaction to capture user perceptions. To permit the examination of these metrics, three different visualizations were created for the nodes interpolation model.

These were:

1. *Interface 1* No visualizations (i.e., an empty 2D display)
2. *Interface 2* Only preset locations displayed
3. *Interface 3* The original nodes interface

These different visual representations for the interface are shown in Fig. 6, Interface 1–3, left to right. In each case, the underlying nodes interpolation model remained the same so that the impact of different visualizations alone could be assessed. This included the layout of the preset sounds within interpolation space and the target location. However, so that this was not obvious to the participants for each interpolator, the interface was rotated through 90° clockwise.

The user testing took the form of a sound design task, where the participants were asked to match a given sound which on the interpolator had a fixed, but unknown to the participants, target location in the interpolation space. A
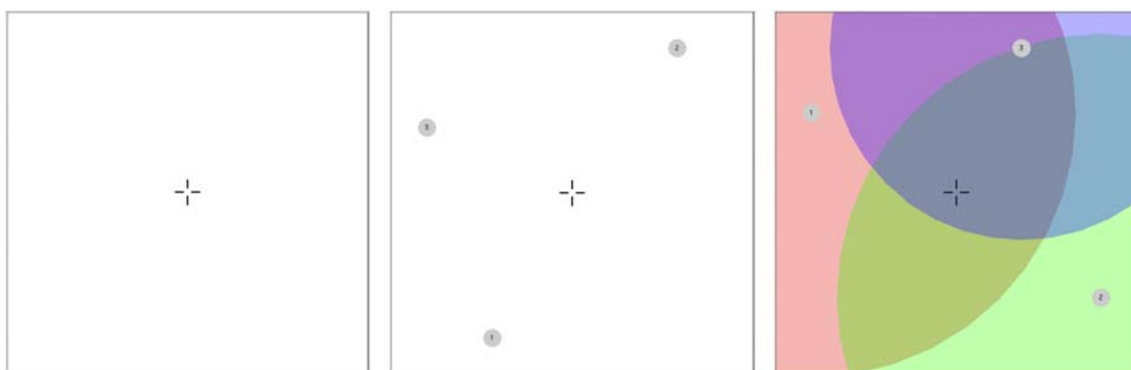
Fig. 6 Different visualizations for nodes interpolator, left to right – no visuals, preset ocations, and full nodes

subtractive synthesis engine (Native Instruments' Massive) was chosen for the experiment as it was found to provide a rich sonic pallet and predictable transitions between sounds. All of the available continuous synthesis parameters (150 in total) were mapped to the interpolators so that every aspect of the preset sounds could be modified. This resulted in an interpolation space with a vast range of distinguishable outputs, with little overlap, that it would be very difficult to explore with Massive's own synthesis user interface. Each interpolator display was populated with different preset sounds, with all of the presets being created from the same base patch, resulting in some sonic commonalities between them. In addition, for each interface, the relative position of the target sound location was the same, meaning that at the target location in each interface generated identical preset weightings.

To simulate a real sound design scenario, the participants were given only three opportunities to hear the target sound before commencing the test. Once the test was initiated, there was no further opportunity to hear the target. In this way, similar to a sound design task, the participants had to retain an idea of the required sound in their "mind's ear." All the participants completed the same sound design task with each graphical interface, but as each interface was set up with different presets, the resulting sonic outputs for each of the three interfaces were different. To minimize any bias through learned experience of using an interpolator, the order in which each participant used the interfaces was randomized. Each test lasted a maximum of ten minutes with the participants being able to stop the test beforehand if they felt the task had been completed. All of the user's interactions with the interfaces were recorded for analysis. When the participants felt that they had matched the required sound they were asked to press a "Target" button so the location could be registered. All other aspects of the interpolation system—inputs, interpolation calculation, mappings (all parameters), and synthesis engine—remained identical between the three interfaces. From this experiment, the raw interaction data could be analyzed to examine differences between the journeys made with the three interfaces.

To assess the perceived usability of the interfaces, the participants were asked to complete a usability questionnaire. The questionnaire was divided into two parts—the first part completed following the use of each interface and the second part filled-in after all three interfaces so they could be compared. The first part utilized one of the most commonly applied and well-tested usability models, the System Usability Scale (SUS) [18]. SUS has been used many times in over a twenty-year period, with a wide range of different systems and has been found to be highly reliable and robust [19], even with small numbers of users. The SUS questionnaire comprises of ten, 5-point Likert items providing a quick assessment of the usability of a system on a scale from 0 to 100. As SUS has been used in many studies, it has been possible to establish

norms which give an indication of a system's perceived usability [20]. It has also been shown that benchmarks for perceived complexity, ease of use, consistency, learnability, and confidence in use can also be extracted from the same survey [21]. The completion of the SUS questionnaire following the use of each interface was to try and establish if the perceived usability changed based on the visuals of the display.

In the second part of the questionnaire, participants were asked which of the three interfaces they preferred and were then asked to rate it on a scale 1–10. To try and understand each participant's level of experience, they were asked how many years they have been using music technology and to rate their sound design experience on a scale 1–10. Finally, they were asked to note down any critical incidents that occur during the test, both positive and negative.

4 Results

The desired number of participants for the experiment was set at fifteen, based on a power assumption of 0.8 and the desire to observe a large effect size (0.5) [22].[1] However, when the experiment was undertaken, sixteen participants were actually recruited, all with some degree of sound design experience. For each participant, their interactions with the interfaces were captured via the recording of mouse movements. This then allowed traces of the movements to be visually compared between the different interfaces. The trace gives a pictorial representation of the journey that each user made through the interpolation space [16]. An example is shown in Fig. 7 for participant 1 who had the following interface order 1, 3, and 2 and participant 2 with interface order 1, 2, and 3. Here they are shown Interface 1–3, left to right with participant 1 on the top row and participant 2 on the bottom.

The cursor mouse position was sampled at a rate of 10 Hz (every 100mS) as it was found to provide enough detail on the cursor movement without overloading the data analysis stage. Also, with this sample rate, it is possible to visually see the relative speed of movements, where slow movements can be seen as smooth lines and fast movements appear as step changes in the trace. In addition, the location of the target sound is shown as a green square and the participants chosen location is indicated by the position of the red square.

It was observed that at the start of the test, while exploring the space, users tended to make large, fast movements. In the middle of the test, the movements tended to slow and become more localized, but a few larger, moderately fast movements were often made. Toward the end of the test, movements tended to slow and become even more focused toward the

---

[1] A minimum of fifteen participants are required to see a large effect size (using Cohen's criteria) with a probability of 0.8 that the effect will be seen in the testing [22].

intended target location. To visualize these aspects, in Fig. 7 the first third of the trace is shown in red, the middle third is shown in blue, and the final third is shown in green. This was also corroborated when the mouse movement speed and distance to target were plotted on a graph, using the same color coding. Figure 8 shows an example for participant 13, with Interface 3.

Broadly these trends were seen in fifteen of the sixteen participants, although as should be expected it did not always evenly divide into thirds of the test time. Nonetheless it appears to indicate that there are three distinct phases during the use of a visual interpolation interface:

1. Fast space exploration to identify areas of sonic interest
2. Localize on "regions of interest", but occasionally check that other areas do not produce sonically better results
3. Refinement and fine tuning in a localized area to find the ideal results

These three phases can be summarized as *exploration*, *localization* and *refinement*. Interestingly these phases were present regardless of the interface being used, showing that the phases must be associated with exploration of the space and not the interface being used [16].

From the traces, it was also observed that as the detail of the visual interface increased so did the distance travelled within the interpolation space. This was despite the fact that the participants were given no information with regard to what the visuals represented. It seems that giving the participants additional visual cues encouraged them to explore those locations. To demonstrate this effect, the mean location for each trace was calculated and the deviation in the form of Standard Distance Deviation [23]. These were calculated with respect to the "unit-square" dimensions of the interface (height and width) and were then plotted back onto the traces to give a visual representation for each interface. An example is shown in Fig. 9 for participant 15 with interface order 1, 3, and 2, although they are shown Interface 1–3, left to right.

It was noted that out of the sixteen participants, thirteen showed an increase in the Standard Distance Deviation when more visual cues were provided by the interface. The normalized standard distances were examined by interface, and it was found for Interface 1 the mean Standard Distance Deviation was 0.131 units ($SD = 0.23$), for Interface 2 the mean Standard Distance Deviation was 0.146 units ($SD = 0.19$), and 0.180 units ($SD = 0.21$) for Interface 3. These results appear to indicate that the presence of more visual cues on the interface tends to encourage wider exploration of the interpolation space, regardless of the fact that the system's output and the goal of the test is sonic.

The locations that the participant actually selected as their target sound were also plotted to see if there were any trends resulting from the different interfaces. Figure 10 shows the selected target locations for all the participants, by interface.
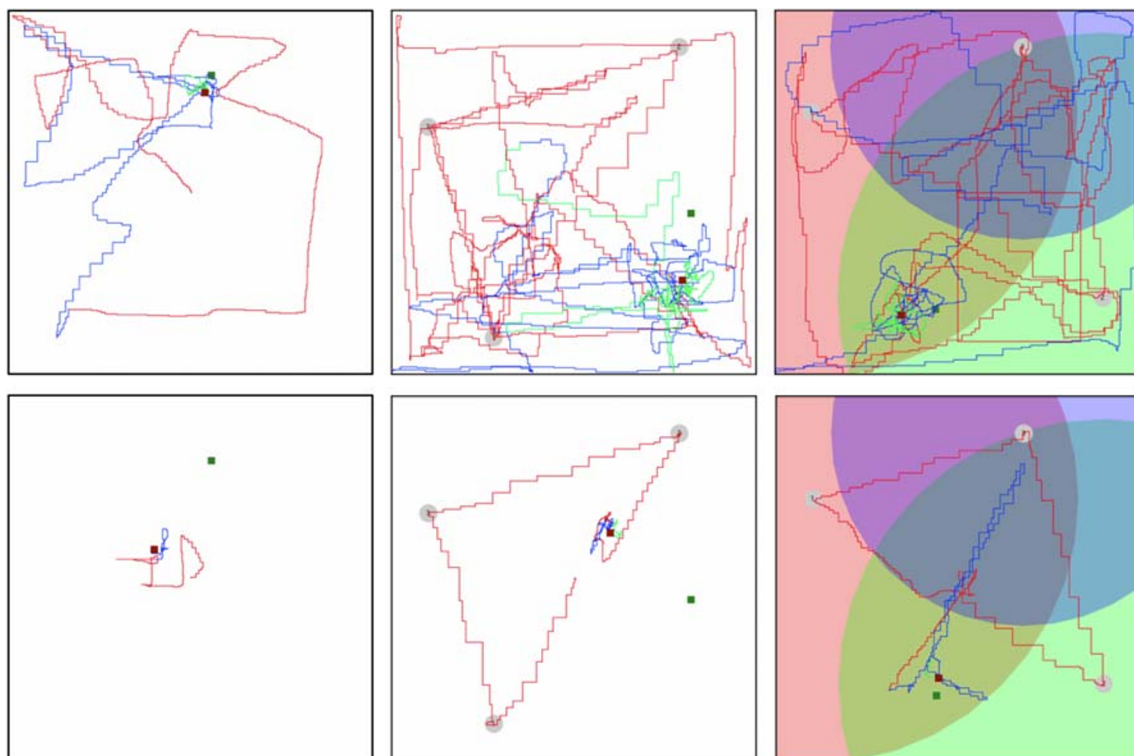
Fig. 7 Mouse traces for participant 1 (top row) and participant 2 (bottom row), showing the location of the target sound (■ green square) and the participants chosen location (■ red square)

It can be seen that with each interface, the users have primarily selected points within the regions of interest (i.e., node intersections), where the sonic output changes. This can be seen as clear localization on the principal axis of variation, albeit rotated 90 degrees for each interface. As this is present with the three different interfaces, it would suggest that the participants used the sonic output to identify the regions of interest.

The results in Fig. 10 show that for Interface 1 (no visualization), there is a fairly wide distribution of locations selected as the target. The Standard Distance Deviation from the correct target location was calculated and found to be 0.300 units for Interface 1 (relative to the height and width). For Interface 2 (preset locations), there is a tighter placement of selected target locations with the Standard Distance Deviation reducing to 0.267 units. Finally, for Interface 3 (full nodes), there is an even tighter localization with the Standard Distance Deviation reducing further to 0.187 units. This appears to show that as the interface provides more detail, it improves users' ability to identify the intended target.

Given that the three interpolators used in the experiment were populated with different preset sounds, they each produced a range of sounds. Therefore, the sonic results for each selected target location were directly compared to the auditioned sound. In all cases, there were sonic differences, but as might be expected, as the selected locations got closer to the true location, the differences became less distinguishable.

4.1 Significance testing

As mentioned previously, there were four areas that it was hypothesized that the visual representation might affect time, speed, distance travelled, and accuracy. Therefore, NHST (Null Hypothesis Significance Testing) was undertaken to establish if there was a significant difference between the interfaces in each of these areas. In all cases, it should be noted that the data was tested for normality (Skewness/Kurtosis, Shapiro-Wilk and visual inspection) and found to lack a normal distribution so non-parametric statistical methods were used. This is most likely due to the exploratory nature of the task being tested. For example, from the start of the test, a participant might immediately move in the right direction, whereas others might move in the opposite direction. Therefore, a Wilcoxon signed-rank test was used to compare the median difference between the interface with no visuals (Interface 1) and the full nodes (Interface 3) visuals [24]. The calculation of effect size for nonparametric statistical methods is less clearly defined than for their parametric equivalents [25]. However, two approaches have become fairly widely adopted for calculating an effect size from the results of a Wilcoxon signed-rank test: correlation coefficient (r) [26] and probability score depth (PS_{Dep}) [27]. For completeness, both of these effect size parameters were calculated from the results obtained.

4.1.1 Time

From the captured mouse data, it was possible to establish the total amount of time that each participant moved the interpolation cursor within the space. It was hypothesized that using the full interface would result in an increase in the cursor movement time over the interface with no visualization (H_A Median$_3$ > Median$_1$). Thus, the null hypothesis was that the different interfaces had no effect on the time the cursor was moved (H_0 Median$_3$ = Median$_1$).

Of the sixteen participants, Interface 3 (full nodes) elicited an increase in the cursor movement time for thirteen

Fig. 8 Mouse distance to target and speed, sampled every 100mS for participant 13 with Interface 3
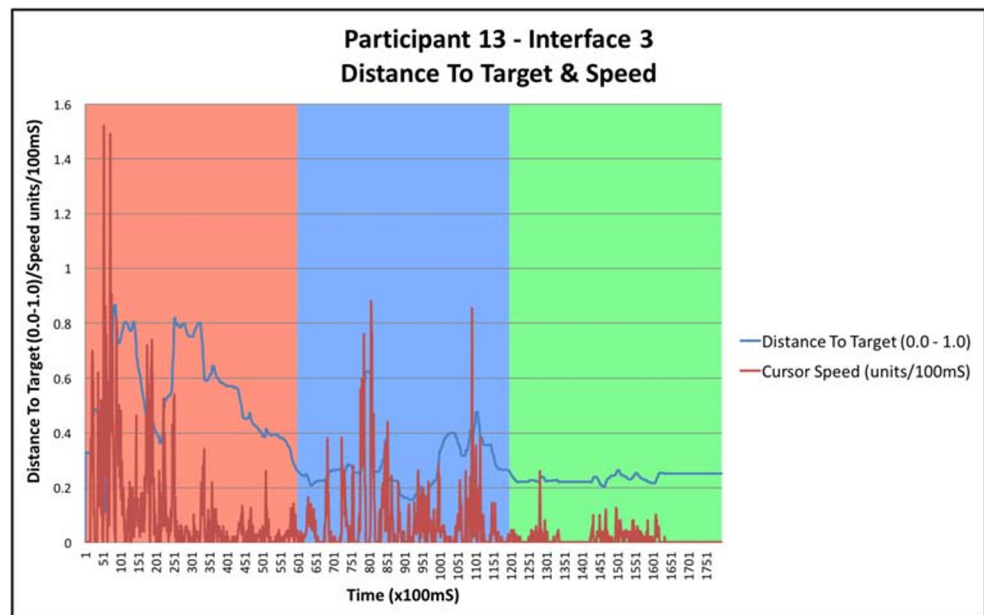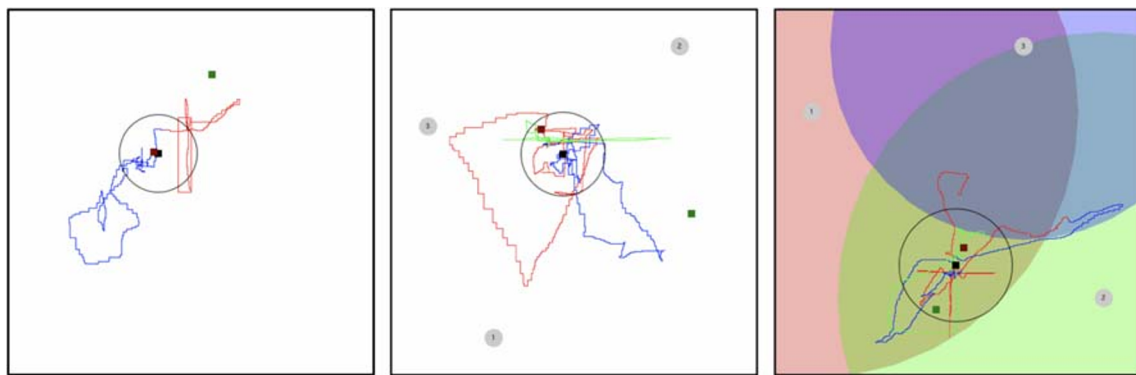
Fig. 9 Mouse trace, mean location (■ black square), standard distance deviation circle, target sound (■ green square), and chosen location (■ red square) for participant 15

participants compared to Interface 1 (no visualization), whereas one participant saw no difference and two participants had a reduced cursor movement time. The difference scores were broadly symmetrically distributed, as assessed by visual inspection of a plotted histogram.

The Wilcoxon signed-rank test determined that there was a statistically significant increase in cursor movement time (Median Difference = 18.0 s, Inter-Quartile Range (IQR) = 56.32–2.85 s) when subjects used Interface 3 (Median$_3$ = 66.5 s, IQR = 129.00–31.72 s) compared to Interface 1 (Median$_1$ = 48.4 s, IQR = 75.32–13.87 s), $Z = -2.669$, $p < 0.008$). The effect sizes ($r = 0.462$, $PS_{Dep} = 0.867$) showed a medium to large effect, where 86.7% of the participants taking the experiment saw an increase in the time they moved the cursor with the full visual interface.

### 4.1.2 Speed, Distance, and Accuracy

A similar methodology was used to compare the median difference for the average cursor speed, the total distance moved by the cursor and the distance from the selected target to the true target location, between Interfaces 1 and 3. Table 1 shows a summary of all the results obtained for the NHST. Note that for all distance-based measures, they are with reference to the

unit-square that is used to plot the visuals within the interface pane (height and width).

From examination of the data in Table 1, it can be seen that with the full interface the participants spent longer exploring, moved faster, and traveled further within the space. It is worth noting that when using the visual interface, the average speed was fairly high at over half the interface width every second. In addition, the interface afforded participants greater accuracy when selecting the target sounds location. As can be seen in all the cases, there is a significant difference between the two interfaces. Using the normal conversions [22], the effect size appears to indicate that in all the cases, it is a medium effect size although a number are approaching the large effect size threshold of 0.5.

4.1.3 Interface 2

The plots and the descriptive statistics generated (shown in Table 2) suggest that Interface 2 generates an intermediate effect between the other interfaces. However, when undertaking NHST, it was not possible to show significance (with 95% confidence interval) between Interface 1 and 2 or Interface 2 and 3. Therefore, these results are considered inconclusive. However, from the descriptive statistics, it appears to be a smaller effect size, and so, it is likely that the small sample
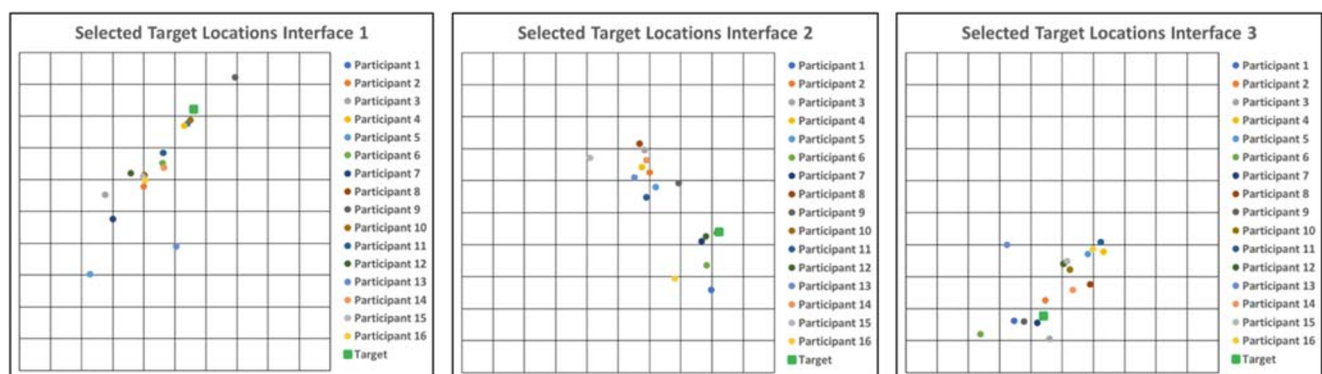
Fig. 10 Participants selected target locations by interface and the location of the target sound (■ green square)

Table 1 Results of the statistical testing to compare participants' performance with different interface visualizations

Test: Wilcoxon signed-rank	Median (IQR)	Significance	Effect size
Time	$Median_3 = 66.5$ s (IQR = 129.0–31.72 s)	$p < 0.008$	$r = 0.462$, $PS_{Dep} = 0.867$
$Z = -2.669$	$Median_1 = 48.4$ s (IQR = 75.3–13.9 s)		
Speed	$Median_3 = 0.569$ units/s (IQR = 0.788–0.283 unit/s)	$p < 0.007$	$r = 0.475$, $PS_{Dep} = 0.937$
$Z = -2.689$	$Median_1 = 0.297$ units/s (IQR = 0.464–0.181 unit/s)		
Distance	$Median_3 = 5.85$ units (IQR = 17.56–3.63 units)	$p < 0.015$	$r = 0.429$, $PS_{Dep} = 0.867$
$Z = -2.430$	$Median_1 = 4.06$ units (IQR = 7.87–1.12 units)		
Accuracy	$Median_3 = 0.177$ units (IQR = 0.248–0.080 units)	$p < 0.039$	$r = 0.366$, $PS_{Dep} = 0.750$
$Z = -2.068$	$Median_1 = 0.264$ units (IQR = 0.367–0.176 units)		

size has not allowed the significance to be shown. Given a larger sample size, it may be possible to show a significance for Interface 2, with the intermediate interface visualizations (preset locations).

4.2 Usability questionnaire

The SUS scores for the three different interfaces were evaluated, and the resulting descriptive statistics are presented in Table 3.

As can be seen from these results, there is very little difference between the mean and standard deviation for the different interfaces. For completeness, a one-way repeated measures ANOVA was performed, which showed that there was no significant difference between the SUS scores for the three different interfaces ($F(2, 30) = 0.378$, $p = 0.688$), given that the sphericity assumption had not been violated ($\chi^2(2) = 1.914$, $p = 0.384$). It appears that the perceived usability of the interpolators is not affected by the visualization. This may be attributed to the fact that the systems functionality and operation is identical with the only change being the visuals presented on the interface. However, in all cases, the SUS scores were extremely high and equivalent to an "A" grade in the 90–95 percentile range, which is equivalent to the top 10% of scores from a database of over 10,000 previous SUS scores from a wide range of computing systems [20, 28], with both simple and complex interfaces [29]. From this database, it has also been shown that the average SUS score is 68, and for all of the interpolators, the mean scores are much higher (82.34%, $SD = 11.68$). To assess if the average SUS scores of the interpolators is significantly different to the

average SUS score, a one-sample t test was performed. This showed that the SUS score was statistically significantly higher by 14.34 (95% CI, 10.90 to 17.79) when compared with the average SUS of 68, $t(47) = 8.367$, $p > 0.0005$. This appears to indicate that the users found the use of interpolators for sound design to be positive. It has been previously suggested that users that give a SUS scores of 82 (± 5) tend to be "Promoters" and likely to recommend the system to other users [28]. Although these norms [20, 28] are from a wide range of general computing systems, the obtained SUS scores still appear to perform well when compared to other music/sound systems that have also been tested with SUS [30–33].

The results for each SUS item were also compared to the defined benchmark values, based on an overall SUS score, in this case the average (68). These have that have been computed based on the data from over 11,000 individual SUS questionnaires and offer a mechanism to examine if specific items have any bias in the overall SUS score [21]. Again, it should be noted that this database consists of results from a wide range of general computer systems, simple and complex. The score for each item in the interface tests were compared to the corresponding benchmark values using one-sample t tests. In all cases, the results show the interpolators perform significantly better than the SUS average, confirming the previous result. The benchmark tests were also repeated for a "good" system with a SUS score of 80. These results did not show significance, so it is not possible to say the interpolators are perceived as being better in some areas of usability than others.

Table 2 Descriptive statistics for Interface 2

Test:	Median (IQR)
Time	$Median_2 = 48.5$ s (IQR = 112.3–30.8 s)
Speed	$Median_2 = 0.374$ units/s (IQR = 0.778 unit–0.233 unit/s)
Distance	$Median_2 = 5.72$ units (IQR = 22.51–3.18 units)
Accuracy	$Median_2 = 0.251$ units (IQR = 0.342–0.130 units)

Table 3 Descriptive statistics for SUS scores by interface

	Interface 1	Interface 2	Interface 3
Mean	82.81	80.62	83.59
Standard error	2.425	3.295	3.251
Median	82.50	82.50	87.50
Standard deviation	9.699	13.182	13.005
C.I. (95%)	5.168	7.024	5.930

In the second part of the questionnaire, participants were asked which of the three interfaces they preferred using. The results are shown in the frequency table in Table 4.

Of the participants, 62.5% preferred using the interpolator with the full nodes visualization, and only 12.5% preferred the interface with no visuals. These results appear to indicate that most users preferred to use interfaces that provide visual cues. To understand the users experience, they were also asked "How many years have you been using music technology." This data was evaluated with respect to their choice of preferred interface giving the results shown in Table 5.

These results appear to show that users that preferred the full nodes interface had an average of 10 years' experience using music technology, as opposed to 7 years and 6.75 years for Interface 1 and 2, respectively. However, given the small sample size between the groups, there is a wide deviation in the means and as such the results should be considered inconclusive. However, when all the participants responses were analyzed together, it did confirm that the participants had a range of different experience levels and there appears to be little bias, although the small sample size (16) should be noted as a larger sample may show other trends. Table 6 shows the results of this analysis.

Analysis of the critical incidents noted by participants seemed to largely mirror the statistical results presented. However, this information did allow an insight into why the participants made the choices they did. The participants that chose Interface 1 (no visualization) made comments that the lack of visuals required them to use their "ears" to locate sounds without decisions being "influenced" by the visuals. Similar remarks, which the sound was primarily used, were also made by those that chose Interface 2 (preset locations), but that the locations were "helpful" and guided them to "pinpointing" sounds. Statements were also made that this interface "matched [their] experience level", but that Interface 3 was "simplest to pick up quickly." Comments made by those that chose Interface 3 (full nodes) related to it "guiding the user how to use it." This appears to be corroborated by multiple participants as comments were made that the visual showed "sound range", "areas that played different sounds", and "where [the sounds] overlap." One participant also stated that the visuals allowed them to "explore sounds without being too lost from [another] sound." It should also be noted that the two the participants that did not choose Interface

3 as their preferred choice stated that the visuals could to be "distracting".

In the general comments, ease of use came up repeatedly, with one participant actually stating that the interpolation systems "did not have a steep learning curve." Another theme that appeared continually was that the participants found the activity "fun" and/or "enjoyable." These aspects appear to support the findings of the SUS and appear to encourage further use of interpolators for sound design activities. In fact, one participant stated they would "like to see it in sound design tools." However, there might be other application areas, as one comment identified that the interpolators maybe useful for "making sounds for EDM (Electronic Dance Music)" and another considered their use for "composition."

5 Discussion

Although the use of sixteen participants in the testing is a small number when compared to usability testing for general computer applications, it is a fair number in the music/sound technology area. Here formal usability testing is not often undertaken, as with the other interpolators [3–10] or where it is done, it is often done with smaller numbers [30–33]. Nonetheless, with sixteen participants, some interesting results were obtained, and significance was shown between how the users interacted with the interface that had no visuals and the one with the full visuals. In addition, to the relatively small number of participants, other limitations of the experiment were that it only tested one specific sound design task and the users were not permitted to change the layout interpolation space. These restrictions were included to constrain

Table 4 Frequency for user preferred interface choice

	Frequency	Percent	Cumulative percent
Interface 1	2	12.5	12.5
Interface 2	4	25	37.5
Interface 3	10	62.5	100
Total	16	100	

Table 5 Descriptive statistics for music technology experience by interface

	Interface 1	Interface 2	Interface 3
Mean	7.00	6.75	10.00
Standard error	1.000	1.250	3.528
Median	7.00	6.50	4.00
Standard deviation	1.414	2.500	11.155
C.I. (95%)	12.71	3.980	7.980

Table 6 Descriptive statistics for music technology experience

	Music technology experience
Mean	8.810
Standard error	2.216
Median	5.500
Standard deviation	8.864
C.I. (95%)	4.730

variability in the experiment, creating consistency between the different interfaces, and focusing the participants on a single activity. This appears to have been successful, and further testing can be undertaken in the future to corroborate the results.

As identified in the Results, the testing showed there appears to be three phases to the identification of a sound with a graphical interpolator system (*exploration*, *localization*, and *refinement*). In the first phase, the users make large, fast moves as they explore the space. During the second phase, the speed tends to reduce as they localize on specific regions of interest. In this phase, though, confirmatory moves have been observed when the user seems to quickly check that there are no other areas that may produce better results. These are inclined to be made at a moderate speed, often in multiple directions. Then in the final phase, the user refines the sound with small, slow movements as they hone in on a desired location. These phases appear to be apparent regardless of the visual display that is presented to the users, with similar phases being observed with all the three of the interfaces tested. However, the frequency of movements, scale, and locations did vary with each participant. This is to be expected as this was an individual task, and the participants possessed different skill levels.

From the examination of all the journeys (mouse traces) for the different interfaces, the visual feedback presented by each affects how the users interact with the systems. When no visualization is provided, the users were effectively moving "blind" and tended to just make random movements within the space. When the preset locations were provided, although the users were not aware of where or how the interpolation was being performed, the provided visual locations encouraged the users to investigate these points and so explore the defining locations. The full interface not only shows the location of the defining sounds but also focuses the exploration and appears to indicate to the users' regions of interest (node intersections for this interpolation model), where there may be interesting sounds. This was also supported by the user's feedback on the questionnaire, where specific comments were made about the identification of overlaps between nodes.

The results from analysis of the journeys appear to be corroborated by the NHST, where we found significant differences in the total time taken to undertake the tests, the average speed of movements during the tests, the total distance moved during the test, and the accuracy in locating a target between the interface with no visuals and the full visual display. Although the primary output from the interpolator is a sonic one, it appears that the feedback provided by the visual display is also of importance. Given the increase in the time taken and distance moved, it appears that the visuals encouraged the participants to explore more of the space and this is also supported by the increase in the standard distance deviation. It also seems that the visual feedback gives users the confidence

to make faster movements with the interpolator. This may be similar to the way a blindfolded person may take longer to explore a space compared to a person without a blindfold, making slower movements and with minimal travel. Finally, given the same activity was being undertaken with each interface, with identical controls, and the goal was a particular sonic output that did not directly relate to the visuals, the increase in accuracy was not foreseen. This could be a secondary effect from exploring more of the space meaning participants were then more likely to locate the correct target. However, it could also be that the full interface provided visual cues so that when a region of interest had been located during the *exploration* phase, the visual cues then made it easier to return to the same area during the *localization* phase. In the same way that a map might aid navigation when trying to discover an unknown location. This was also highlighted by one of the participants who said that the visual stopped them getting lost after finding a sound of interest.

Interestingly, although the visual representation was not explained to the participants, it is clear from the participant feedback that they were able to work out from the visuals that what they meant and their function in terms of the system's operations. This appears to imply that the system with the full visuals is intuitive and guides the user in its operation.

Given that no significant difference in the SUS scores was observed between the different interfaces, it may be said that there is no difference in the perceived usability, despite the fact that subjects' behavior differed between Interfaces 1 and 3. However, as the overall average SUS score is so high, it appears to suggest that just the concept of using a graphical pane to control interpolations is considered highly usable in itself. This seems to be supported by the general comments made by the participants that they enjoyed using the interpolators and found the experience to be fun. Moreover, given that there seems to be little difference in the perceived usability, based on the visualization, and some participants stated they preferred fewer visual cues, there is a case for giving the user control of the level of detail provided by the interface. In this way, users that just want to concentrate on the systems sonic output can do so with no visual guidance, and those that find the visual useful could customize the level of detail for their particular needs or preference.

6 Conclusions

The identification of the three distinct phases of use during the testing of the graphical interpolators is of significant interest as it suggests that users interact with the interfaces differently at different stages during their journey through the interpolation space. Better understanding of the user behavior with these systems will allow further evaluation of different interface visualizations. Moreover, in future work this information

could be used in the design of new interfaces that provide users with visuals that facilitate the different phases of a sound design task. In addition, there is no reason why the visuals have to remain static and maybe they should change as the users enter the different phases of the process using an Interactive Visualization (IV) paradigm [34]. This could either be through allowing the user to select different visualization or based on their interactions with the space.

From the results obtained it appears that the visual display of an interpolator's interface has a significant impact on the sonic outputs obtained. The resulting journeys made with the full interface show a wider exploration of sonic outputs, faster speed of movement, and improved accuracy at locating a specific sound. The effect sizes are relatively large, giving greater confidence in the validity of the results. At the same time, however, it appears the system is still perceived by the users as highly usable and unaffected by the change in visualization. This again adds further strength to the idea of using IVs that can be adapted to the user needs.

A number of different visual models have been previously presented for graphical interpolators [3–10], each of these using very different visualizations. In previous work, it has been shown that these different interpolation models generate very different sonic palettes, even when populated with identical sounds [11]. Given now the suggested importance of the visual feedback provided by each interface, it will be important in future work to evaluate the suitability and relative merits of each through further user testing. Based on the results from this study, there are two areas to be refined. Although for this experiment, the directive nature of task, asking all the participants to locate the same sound, worked well as it allowed direct comparisons between the different levels of visualization, it is perhaps not truly representative of a real sound design task. Sound design is a highly creative and individual practice so typically there are likely to be different choices made between designers. As a result, in the testing of the different interpolators, a more typical sound design scenario will be adopted. It is also anticipated that this will help to identify not only the positives of using interpolated interfaces for sound design but also the potential limitations.

Another area that will be examined further is the use of the standard SUS questionnaire to compare the usability. While SUS showed that this interpolator is considered highly usable, it did not give enough detail to show any perceived differences between the visualizations, even though the quantitative data suggests otherwise. For comparison purposes, SUS will be used in subsequent interpolator testing, but to probe this area further, the standard SUS questions will be augmented with additional items, more specific to the interface. This strategy has been shown to work successfully in other evaluations of audio technology interfaces [35].

Finally, given that it appears graphical interpolators are perceived as highly usable in this application area, further consideration should perhaps be given to their wider use in other domains. This was even highlighted by participants taking part in this study who made suggestions relevant to their own practice, as detailed in the Usability questionnaire. In the music/sound area, they could be further utilized for generation, composition, performance, or musical expression, as well as sound design, while in a wider context, graphical interpolators could be beneficial for the control of graphics, animation, texturing, image-processing, database transactions, avatar generation, game-level design, etc. In fact, graphical interpolation lends itself to any situation where new states require exploration and/or identification, based on a set of known states, particularly within dense parameter spaces.

Compliance with ethical standards

References

1. Hunt A, Wanderley M, Paradis M (2003) The importance of parameter mapping in electronic instrument design. J New Music Res 32(4):429–440. https://doi.org/10.1076/jnmr.32.4.429.18853
2. Van Nort D, Wanderley M, Depalle P (2014) Mapping control structures for sound synthesis: functional and topological perspectives. Comput Music J 38(3):6–22. https://doi.org/10.1162/COMJ_a_00253
3. van Wijk JJ, van Overveld CW (2003) Preset based interaction with high dimensional parameter spaces. In: data visualization 2003. Springer US, pp 391–406. https://doi.org/10.1007/978-1-4615-1177-9_27
4. Allouis JF (1982) The SYTER project: sound processor design and software overview. In: Proceedings of the 1982 International Computer Music Conference (ICMC). Michigan Publishing, University of Michigan Library, Ann Arbor
5. Spain M, Polfreman R (2001) Interpolator: a two-dimensional graphical interpolation system for the simultaneous control of

digital signal processing parameters. Org Sound 6(02):147–151 https://doi.org/10.1017/S1355771801002114

6. Bencina R (2005) The metasurface: applying natural neighbour interpolation to two-to-many mapping. In Proceedings of the 2005 conference on New interfaces for musical expression, May 1 (pp. 101–104). National University of Singapore

7. Larkin O (2007) INT.LIB–a graphical preset interpolator for Max MSP. ICMC'07: In Proceedings of the International Computer Music Conference, 2007

8. nodes (2019) Max Reference, Cycling 74

9. Drioli C, Polotti P, Rocchesso D, Delle Monache S, Adiloglu K, Annies R, Obermayer K (2009) Auditory representations as landmarks in the sound design space. In Proc. of Sound and Music Computing Conference

10. Marier M (2012) Designing mappings for musical interfaces using preset interpolation. In Conf. on New Interfaces for Musical Expression (NIME 12)

11. Gibson D, Polfreman R (2019) A Framework for the Development and Evaluation of Graphical Interpolation for Synthesizer Parameter Mappings In Proceedings of Sound and Music Computing Conference, 2019. https://doi.org/10.5281/zenodo. 3249366

12. Van Nort D, Wanderley M (2006) The LoM mapping toolbox for Max/MSP/Jitter. In Proc. of the 2006 International Computer Music Conference (ICMC06) (pp. 397–400)

13. Bevilacqua F, Müller R, Schnell N(2005) MnM: a max/MSP mapping toolbox. In proceedings of the 2005 conference on new interfaces for musical expression (pp. 85-88). National University of Singapore

14. Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: a review. Environ Model Softw 53:173–189

15. Sarter NB (2006) Multimodal information presentation: design guidance and research challenges. Int J Ind Ergon 36(5):439–445

16. Gibson D, Polfreman R (2019) A journey in (interpolated) sound: impact of different visualizations in graphical interpolators. In Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound (pp. 215–218)

17. International Organization for Standardization (2018) Ergonomics of human-system interaction — part 11: usability: definitions and concepts (ISO 9241-11:2018). https://www.iso.org/standard/63500. html

18. Brooke J (1996) SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL (eds) Usability evaluation in industry. Taylor and Francis, London

19. Bangor A, Kortum P, Miller JA (2008) The system usability scale (SUS): an empirical evaluation. Int J Human-Compt Interact 24(6): 574–594. https://doi.org/10.1080/10447310802205776

20. Brooke J (2013) SUS: a retrospective. J Usability Stud 8(2):29–40

21. Lewis JR, Sauro J (2018) Item benchmarks for the system usability scale. J Usability Stud 13(3):158–167

22. Cohen J (1988) Statistical power analysis for the behavioral sciences. Routledge Academic, New York

23. Mitchell A (2005) The ESRI guide to GIS analysis, volume 2: spatial measurements & statistics. ESRI Press, Redlands

24. Pallant J (2016) SPSS survival manual: a step by step guide to data analysis using SPSS for windows, 6th edn. McGraw Hill Open University Press, New York

25. Leech NL, Onwuegbuzie AJ (2002) A call for greater use of non-parametric statistics. In Mid-South Educational Research Association Annual Meeting

26. Rosenthal R (1994) Parametric measures of effect size. In: Cooper H, Hedges LV (eds) The handbook of research synthesis. Russell Sage Foundation, New York, pp 231–244

27. Grissom RJ, Kim JJ (2012) Effect sizes for research: Univariate and multivariate applications, 2nd edn. Taylor & Francis, New York

28. Sauro J (2011) A practical guide to the system usability scale: background, benchmarks, & best practices. Measuring Usability LLC, Denver

29. Kortum PT, Bangor A (2013) Usability ratings for everyday products measured with the system usability scale. Int J Human-Compt Interact 29(2):67–76

30. Yang YC, Wang ST, Tseng YJ, Lin HCK (2010) light up! Creating an interactive digital artwork based on Arduino and Max/MSP design. In 2010 International Computer Symposium (ICS2010) (pp. 258-263). IEEE. https://doi.org/10.1109/COMPSYM.2010. 5685506

31. Order S (2015) 'ICreate': preliminary usability testing of apps for the music technology classroom. J Univ Teach Learn Pract 12(4):8

32. Spiliotopoulos K, Rigou M, Sirmakessis S (2018) A comparative study of skeuomorphic and flat design from a UX perspective. Multimodal Technol Interact 2(2):31. https://doi.org/10.3390/ mti2020031

33. Stolfi AS, Milo A, Barthet M (2019) Playsound. space: improvising in the browser with semantic sound objects. J New Music Res 48(4):366–384. https://doi.org/10.1080/09298215.2019.1649433

34. Keefe DF, Isenberg T (2013) Reimagining the scientific visualization interaction paradigm. Computer 46(5):51–57. https://doi.org/ 10.1109/MC.2013.178

35. Wilson A, Fazenda BM (2019) User-guided rendering of audio objects using an interactive genetic algorithm. J Audio Eng Soc 67(7/8):522–530