HAL
archives-ouvertes.fr

Combining audiovisual mappings for 3D musical interaction

Florent Berthaut, Myriam Desainte-Catherine, Martin Hachet

▶ **To cite this version:**

Florent Berthaut, Myriam Desainte-Catherine, Martin Hachet. Combining audiovisual mappings for 3D musical interaction. International Computer Music Conference, Jun 2010, New york, United States. p 100, 2010. <hal-00530076>

HAL Id: hal-00530076
https://hal.archives-ouvertes.fr/hal-00530076

Submitted on 27 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COMBINING AUDIOVISUAL MAPPINGS FOR 3D MUSICAL INTERACTION

Florent Berthaut
University of Bordeaux
SCRIME
LaBRI

Myriam Desainte-Catherine
University of Bordeaux
SCRIME
LaBRI

Martin Hachet
INRIA
LaBRI

ABSTRACT

3D environments provide new possibilities for musical interaction. They allow musicians to manipulate and visualize large sets of sound processes associated to 3D objects by connecting graphical parameters to sound parameters. Several of these audiovisual mappings can be combined on a single 3D object. However, this brings up the issues of the choice of these mappings and of their combinations. We conducted a user study on sixteen musicians to evaluate audiovisual mappings and their combinations in the context of 3D musical interaction. This user study is composed of three experiments. The first experiment investigates subjects preferences for mappings between four perceptual sound parameters (amplitude, pitch, spectral centroid and noisiness) and ten graphical parameters, some of them specific to 3D environments. The second experiment focuses on efficiency of single mappings in an audiovisual identification task. The results show almost no significant differences, but some tendencies, which may indicate that the choice of mappings scales is more important than the choice of the mappings themselves. The third experiment investigates the efficiency of mappings combinations. The results indicate no significant differences, which suggest that it may be possible to combine up to four audiovisual mappings on a single graphical object without any performance loss for musicians, if they do not disrupt each other.

1. INTRODUCTION

The past decade has seen an important development of graphical musical interfaces. These interfaces don't share the physical limitations of hardware controllers. Consequently, they give access to a high number of control parameters of an unlimited number of sound processes. Each musical parameter can be modified by interacting with the graphical components of the interface. Furthermore, in opposition to traditional instruments, graphical interfaces allow users to display precise information about the musical events, such as their activity, their current parameters and so on. Immersive virtual environments, used in the field of virtual reality, add even more possibilities for musical applications. In particular, they enable the visualization of large sets of audiovisual objects, with additional parameters such as distance, 3D orientation, 3D shape, transparency. 3D interaction techniques, such as virtual sculpting, ray-casting, or navigation, may also be favourably used for musical purposes. These interfaces bring up the issue of the choice of audiovisual mappings.

Which graphical parameters of 3D objects should be used to display sound parameters for visualization and manipulation ? Appropriate mappings and visual organisation may improve interaction efficiency as it is done in the information visualization field. Moreover, using combinations of audiovisual mappings, as is it done for generic data by Healey [6], users may manipulate and visualize many musical parameters on a single 3D object, in opposition to simple sliders controlling only one parameter each. Thus it is important to evaluate these different mappings, in terms of subjects preferences and performances, in the context of musical interaction, i.e with varying sound parameters.

In this paper, we present a user study that investigates user rates, in section 3.5, and efficiency, in section 3.6, of audiovisual mappings for 3D graphical musical interaction. It is also aimed at evaluating the quantity of musical information, i.e sound parameters, that a graphical component may display without causing disruption. This last evaluation is done in the third experiment, described in section 3.7.

2. RELATED WORK

A lot of work has been done on linking images to sound or music [10], from Isaac Newton who noticed the correspon-

dence between prismatic rays and musical scales, to the first color-music instruments such as "Le clavecin oculaire" from Louis-Bertrand Castel, to the extensive work of John Whitney [15].

Existing 3D instruments rely on traditional visualization of sounds, such as 3D spectrograms or 3D speakers for spatialization [1]. None of them combine several audiovisual mappings on the 3D objects, thus they reduce the visualization and interaction possibilities. For example, 3D spectrograms must be viewed from a certain angle to be understandable, so that users cannot see or manipulate many sounds processes representations at the same time. Other applications use simple sliders transposed to 3D, reducing the control possibilities brought by 3D objects to a single parameter.

Many studies were done on audiovisual mappings. Walker [14] studied extensively both musicians and non-musicians preferences concerning audiovisual mappings, with a musical notation perspective, using sequences of static 2D graphical signs matched with sound sequences. He found significant preferences, but only within the musicians groups. Another study was made by Lipscomb [11], in order to build a visualization tool. Variations between six distinct instruments timbres were matched to shape variations. In opposition to previous studies, no significant differences were found between musicians and non-musicians. According to the author, this could be related to the low number of subjects. Visual stimuli were three states animations of 2D shapes. Finally, very interesting work was done by Giannakis [5, 4, 3], that led to Sound Mosaic, a tool for synthesis and composition. He concentrated on perceptual dimensions of color and texture as visual parameters, thus following research of the information visualization field. Subjects had to choose sequences of images from palettes to represent sequences of sounds consisting of linear or random variations of the auditory parameters. The results of these studies are presented in figure 1.

However these user studies do not answer our specific questioning, which is the audiovisual mappings of dynamic parameters on 3D objects and their combinations.

Author	Auditory Feature	Visual Feature
Walker[14]	loudness	size
	pitch	vertical position
	timbre	pattern
	duration	horizontal length
Lipscomb[11]	loudness	size
	loudness	color hue
	pitch	vertical position
	pitch	color hue
	timbre	shape
Giannakis[5]	loudness	color saturation
	pitch	color brightness
	dissonance	text. repetitiveness
	sharpness	text. coarseness
	compactness	text. granularity

Figure 1. Results of Previous User Studies for AudioVisual Mappings

3. USER STUDY

3.1. Overview

This study tried to answer three questions. Do mappings preferences observed in other studies still appear with 3D representations ? Are some audiovisual mappings more efficient for an identification task than others, especially in a dynamic context ? Does combining several mappings on a single graphical component improve performances of an identification task ?

In the first experiment, we studied the preferences of the subjects on the mappings between four audio parameters and ten graphical parameters. These first results were used to reduce the number of mappings, keeping the preferred ones. In the second experiment, we studied the efficiency of these mappings. To do so, we measured the subjects' performances in an audiovisual object identification task. The results were then used to choose one graphical parameter for each audio parameter. In the third experiment, we studied the effect of combining the chosen mappings on the subjects' performances.

3.2. Stimuli

3.2.1. Auditory stimuli

Several types of sound parameters could be controlled and visualized. Perceptual parameters are for example loudness, pitch, and the different dimensions of timbre. Low-level physical parameters include the waveform and the spectrum. There are also parameters of the different sound synthesis techniques, such as additive, substractive, granular and so on. Finally there are sound processing parameters, including the parameters of effects, for example pitchshifting, distortion and filter.

Perceptual parameters provide numerous advantages. They are more general because they can be extracted from any synthesized or recorded sound source, more easily than sound processing parameters. Moreover, they are understandable, as opposed to some synthesis parameters, so that one clearly perceives the effect of their variations. Some of them can be modified by a single standard audio effect, such as pitchshifting, volume, distortion or low-pass filter. Thus they can be controlled as well as visualized.

Four auditory perceptual parameters were used in the experiments: pitch, amplitude/loudness, brightness / spectral centroid and noisiness / irregularity. Seven audio sequences of 7 seconds were generated for each parameter using the Renoise[1] tracker, starting with a simple sinusoid and then using Renoise's internal effects to produce variations. The audio loops were played and analysed in realtime to get the perceived audio parameters in PureData[2].

[1] http://www.renoise.com/
[2] http://puredata.info/

The amplitude parameter featured random discrete and continuous variations ranging from 0dB to 30dB and analysed with the env~ object. The discrete variations of the pitch parameter were values randomly chosen within the notes {60, 63, 67, 68, 72}, the corresponding frequencies being {261.6Hz, 311.1Hz, 391.9Hz, 415.3Hz, 523.2Hz}, and analysed with the fiddle~ external [13]. The pitch variations were small (within one octave) in order to prevent variations of perceived loudness. The brightness / spectral centroid changed with random discrete and continuous variations ranging from 948Hz to 998Hz and analysed with the sc~ external from the flib library. Finally, the noisiness / irregularity had random discrete and continuous variations ranging from 3.6 to 170 and analysed with the irreg~ external from the flib library, calculated according to Jensen [8] and based on the original formula by Krimphoff [9].

Seven sequences of 7 seconds were also made by combining variations of the four parameters.

3.2.2. Visual stimuli

In order to provide fast and precise graphical control and visualization, perceptually salient parameters are needed. The field of information visualization has proven the efficiency of preattentive visual features, as described by Healey [7]. These visual properties are detected without the need for focused attention, independently from the number of displayed elements, so that tasks involving them, such as detecting a filled circle in a group of empty circles, take 200ms or less. Target detection or estimation of the number or percentage of visual elements can thus be improved by using them. Examples of preattentive features are 2D/3D orientation, length, size, curvature, number, color hue, intensity, direction of motion, stereoscopic depth, lighting direction. Some of these features can be organized into a hierarchy, for example color luminance affects color hue perception which in turn affects visual texture perception. As explained in the introduction, these features are also often combined, for example in [6], to enable efficient visualization of complex data. Furthermore, studies by Luck and Vogel [12] demonstrated that it is possible to retain information about four visual objects defined by a conjunction of four features in working memory.

For our experiment, ten graphical parameters were chosen, some of them being preattentive features. The graphical parameters were applied to cubes rendered in a 3D environment using the OpenSG[3] scenegraph library. Distance on the z-axis ranged from the positions -1 to -3, camera being on 0. Ground and shadows were also used in order to improve its perception. Translations on x and y axis were not tested, in order to save them fo the spatial organisation of the 3D objects. The orientation parameter was a rotation around the z-axis ranging from 0rad to 1rad, so that the rotation was understandable. The size parameter was a scale operation ranging from 0.4 to 0.9, so that the object did not disappear. The texture rugosity parameter was rendered using relief bump mapping with scale parameter ranging from 0 (flat) to 10 (very rough). The scale of the object's texture varied for the texture scale parameter from original size to eight times this size. The speed parameter was rendered by a circular motion ranging from 0 to Pi unit per second. One dimension only was selected for the color parameter, color lightness, mostly because we wanted to keep the experiments short enough. There is a hierarchy between features, as said before luminance perturbs hue perception, and that would have cause problems in the combinations experiments. The color was ranging from {0,0,0} to {255,255,255}. The shininess parameter ranged from 0 to 1. It corresponded to the amount of specular reflection of the 3D objects surfaces. The shape distortion parameter corresponded to the scale of a random modification of the object's vertices' positions, ranging from 0 to 0.5 . Simple generative shapes [2] or more complex shapes could have been experimented, but it was instead decided to choose a parameter which could affect any complex shape. Indeed, because complex shapes allow for multiple dimensions whithout disturbing other features, they could be used to represent multidimensional perceptual phenomena, such as the sound spectrum divided into the twenty-four critical bands of hearing [16], and thus they were not selected for this experiment. The transparency parameter varied from 0 to 0.8, so that the object did not completely disappear. To make sure that transparency was correctly perceived, a randomly generated background was added.

The 3D animations were rendered in realtime, visual parameters being connected to the audio parameters by OpenSoundcontrol [4] messages which were sent from PureData to the 3D software. The scales of the graphical parameters were linear and chosen in order to fit the ranges of the audio parameters.

3.3. Subjects

There were sixteen subjects (13 males and 3 females) aged between 22 and 54. All of them were trained musicians (having at least taken music lessons), and they all had already used music software. None of them were regular users of 3D software.

3.4. Experimental Setup

Subjects sat in front of a 12,1" laptop screen and they were equipped with Beyerdynamic DT-770 headphones. They went through the experiments and entered their answers using a modified computer keyboard on which all the keys

[3] http://opensg.vrsource.org/trac

[4] http://opensoundcontrol.org/

were removed, except the spacebar and the four keys used to select the objects in the second and third experiments.

3.5. First experiment: Single mappings - Preferences

3.5.1. Goals

The goal of this first experiment was to test subjects preferences for all the possible mappings. Our hypothesis was that several significant preferences would appear, similar to what has been found in existing user study.

3.5.2. Experimental procedure

This experiment was composed of four tests, one for each sound parameter, i.e amplitude, pitch, centroid and noisiness. Ten cubes were displayed in each test, each cube using a different graphical parameter to display the sound parameter variations, as seen on figure 2. The subjects were asked to rate verbally the relevance of each mapping between 1 and 5. The order of the tests was randomly chosen.

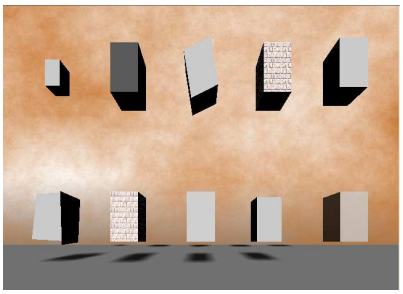

Figure 2. Screenshot of the first experiment

3.5.3. Results and analysis

The results are shown in figures 3, 4 and 5. Kruskal-Wallis tests indicated a significant effect for Amplitude ($\chi^2(9) = 57.81, p < 0.0001$), Pitch ($\chi^2(9) = 32.178, p = 0.0002$), Spectral Centroid ($\chi^2(9) = 29.932, p = 0.0005$), but no significant effect for Noisiness. Wilcoxon analysis for any two mappings of the significant tests revealed the significant differences shown below each figure.

3.5.4. Subjects comments

Some subjects would have preferred an inverted scale for some mappings. For example, they felt that higher pitch should correspond to smaller object. This is consistent with the the fact that high frequencies are perceived louder than low frequencies. Most subjects assimilated distance to a 1D fader. They also reported that changes in the sound identity, ie. timbre parameters, should be associated to changes in

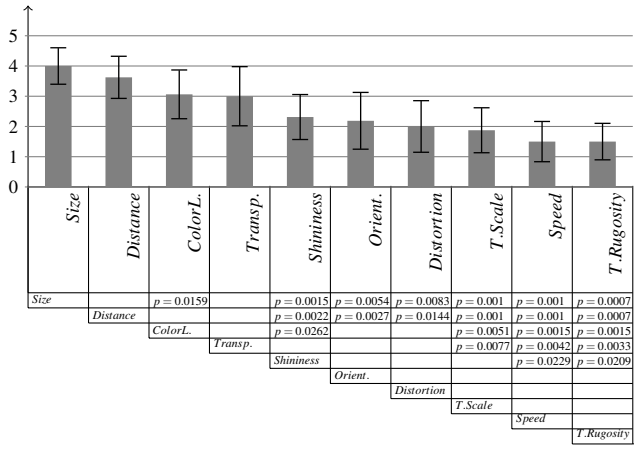

Figure 3. First experiment : Preferences - Subjects rates for Amplitude (and average deviations). The significant differences between the mappings from a Wilcoxon analysis are also given.

the object's identity, i.e shape. Subjects felt that the speed parameter should have been applied to something else than a circular movement.

3.5.5. Discussion

The results indicated significantly preferred graphical parameters for amplitude, pitch and spectral centroid. In particular, subjects associated amplitude with size as in Walker's experiments, but also with distance and transparency. The tendency for the pitch parameter was to associate it with distance, which subjects in their comments described as a 1D fader. Thus it seemed to correspond to the vertical position used in Walker's study. Color lightness was significantly preferred for spectral centroid and the tendency for noisiness was to associate it with texture rugosity. That could be explained by the correlation seen in the other studies between the inner aspect and the timbre parameters. We chose to remove the graphical parameters with the worst ranks for these four audio parameters and keep the five best ones, to evaluate the mappings with our performance experiment.

3.6. Second experiment: Single mappings - Performances

3.6.1. Goals

The goal of this second experiment was to test, for each sound parameter, the efficiency of the five graphical parameters that had been identified as the preferred ones in the first experiment. This was done by measuring the subjects performances in an identification task. For the amplitude parameter, size, distance, color, transparency and shininess were tested. For the pitch parameter, distance, color lightness, shininess, size and orientation were tested. For the spectral centroid parameter, the tested parameters were color,

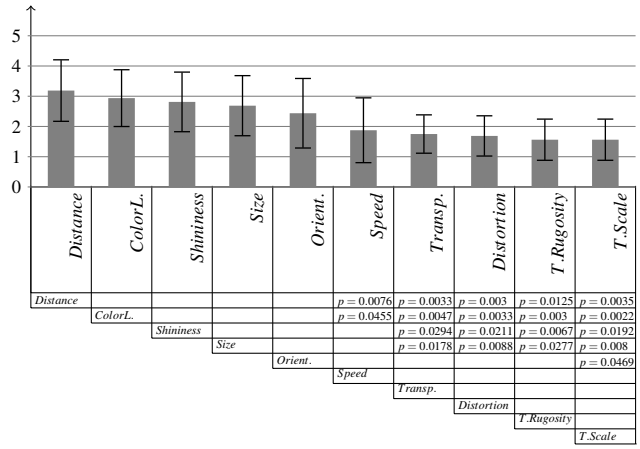

Figure 4. First experiment : Preferences - Subjects Rates for Pitch (and average deviations). The significant differences between the mappings from a Wilcoxon analysis are also given.

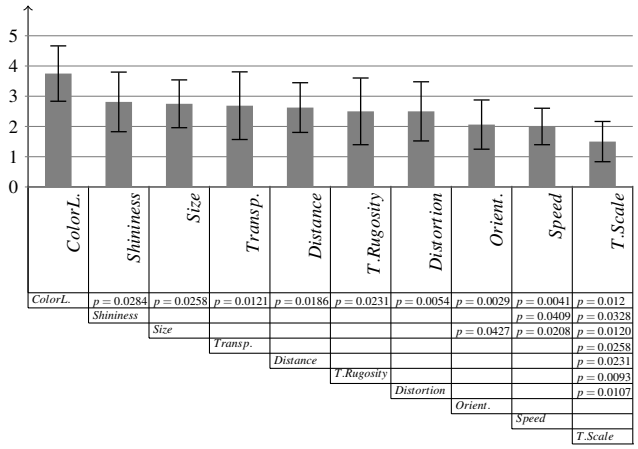

Figure 5. First experiment : Preferences - Subjects rates for Spectral centroid (and average deviations). The significant differences between the mappings from a Wilcoxon analysis are also given.

shininess, size, transparency and distortion. And for the noisiness parameter, texture rugosity, color, distortion, size and transparency were tested.

Our hypothesis was that some graphical parameters would be more efficient than others, because musicians were used to some representations.

3.6.2. Experimental procedure

The second experiment was composed of four tests, one for each sound parameter. Each test involved 5 conditions, one for each selected graphical parameter, and each condition involved seven trials. During each trial, an audio sequence corresponding to the sound parameter was played and four cubes were displayed, as seen in figure 6. Subjects were asked to identify as quickly as possible the cube whose graphical variations were connected to the sound parameter. They selected it with the corresponding key of the keyboard. Response times and error rates were recorded. At the end of each condition, subjects were also asked to give an evaluation of its easyness between 1 and 5. To avoid a learning effect, the order of the sequences and of the conditions for each audio parameter were randomly chosen. A practice trial was also added at the beginning of each condition.

3.6.3. Results

The response times are shown in figure 7. An analysis of variance (ANOVA) indicated a significant effect of the mapping choice for Amplitude ($F = 6.131 > F(4,60,0.05) = 2.525$), Pitch ($F = 4.132 > F(4,60,0.05) = 2.525$), and Noisiness ($F = 4.025 > F(4,60,0.05) = 2.525$), but no significant effect for Spectral Centroid. A Student-Neumann-Keuls for any two mappings of the significant tests revealed

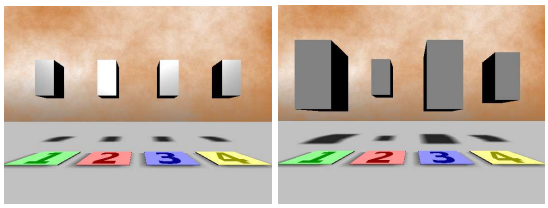

Figure 6. Second experiment screenshots: Amplitude mapped to Shininess and Amplitude mapped to Size

the following significant differences. For the Amplitude parameter, Size, Distance, Transparency and Color Lightness were more efficient than Shininess. For the Pitch parameter, Distance was more efficient than Orientation, and Distance and Orientation were more efficient than Shininess. For the Noisiness parameter, Size was more efficient than Texture Rugosity.

An ANOVA indicated no significant differences for the error rates of the four tests.

A test of correlation was done for each user between their preferences in the first experiment and their response times of the performances experiment. It revealed that there was no correlation between these results, with an average value of 0.16.

3.6.4. Subjects comments

Subjects preferred mappings that had an analogy with the real world, for example when sound got brighter as the object got closer, when sound got louder when the object got bigger or when sound got higher pitched when the object got

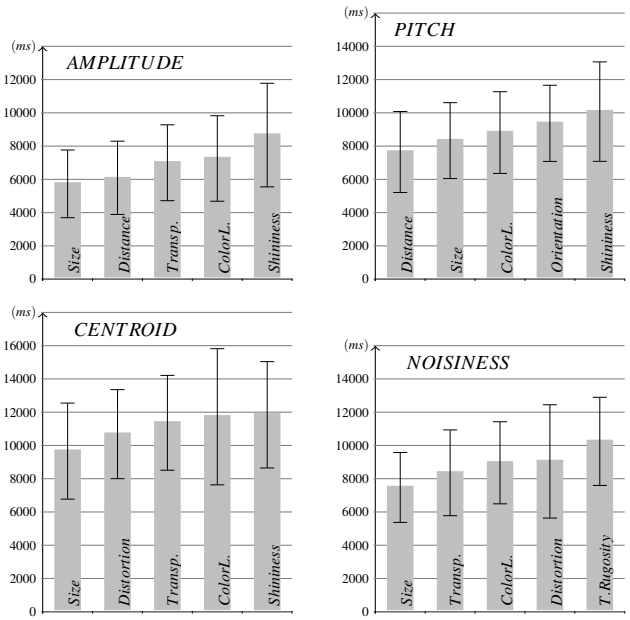

Figure 7. Second experiment : Performances - Response times (in ms) (and average deviations) for Amplitude, Pitch, Spectral centroid and Noisiness

smaller. They indicated that the scale for the orientation parameter was not well chosen, and also that the changes in the rugosity parameter was not correctly discernible. Subjects felt that the scale of the transparency should be modified to better fit some of the audio parameters. They also expressed the need of reversing the scale for some mappings. Subjects interestingly reported that they used mainly quick variations of the parameters to identify the correct object, because they perceived them better than slow variations, no matter which graphical parameter was tested.

3.6.5. Discussion

First of all, our results indicated that there was no correlation between the preferences and the performances of the subjects. In other words, the preference of a user for one mapping did not mean that this mapping was the most efficient. This confirmed the need for a user study with objective performance measurements.

We almost did not find any significant differences between response times for the different mappings, while we had found some in the preferences experiment. Differences that were found for shininess, orientation, and texture rugosity, seemed to be scale problems. That was indeed confirmed by the subjects comments. Moreover, subjects reported that they followed fast variations, no matter which graphical parameter was used, to identify which object was associated to the sound. This led us to think that the scales of the mappings were finally more important than the mappings themselves.

Even if the precedent results had shown no significant differences between the mappings, we chose by a process of elimination which graphical parameter we would use for each audio parameter, in relation to the tendencies revealed by the results. So we finally chose the following mappings: Size with Amplitudes, Color with Pitch, Distortion with Spectral Centroid and Transparency with Noisiness.

3.7. Third experiment: Combined mappings

3.7.1. Goals

The goal of this last experiment was to test the influence of the number of simultaneous audiovisual mappings on performances. Our hypothesis was that a larger number of audiovisual mappings would improve the identification task performances, because subjects would have several perceptual clues at the same time.

3.7.2. Experimental procedure

The third experiment was made up of two parts, each with four sequences and seven trials by sequence. Subjects again had to choose which of the four cubes displayed was connected to the sound they were hearing by selecting it with the corresponding key of the keyboard, and then give a score for each sequence. For the first part, the sound was a simple sinusoid with only amplitude variations. In the first sequence of this part, only the amplitude-size mapping was activated. In the second, third and fourth sequences, respectively one, two and three additional graphical parameters were randomly animated, to disrupt the subject. For the second part, the sound was a sinusoid with combined variations of the four audio parameters. In the first sequence, only one mapping was activated, size was connected to the amplitude. In the second, third and fourth sequences the following mappings were respectively added: color and pitch, distortion and brightness, transparency and noisiness, as shown in figure 8. In order to avoid a learning effect, the order of the sequences in each part was randomly chosen. This experiment was ran a second time, replacing the color parameter with the distance parameter, to test the interference between size and distance variations.

3.7.3. Results

The response times of the first part of the third experiment are shown in figure 9. An ANOVA indicated a significant effect ($F = 9.498 > F(3,45,0.05) = 2.812$) of the number of disrupting graphical parameters when using the distance parameter. Student-Neumann-Keuls analysis revealed a significant difference between the first sequence and the three other sequences with disruptions.

When using the color parameter, the ANOVA also indicated a significant effect ($F = 10.094 > F(3,45,0.05) =$

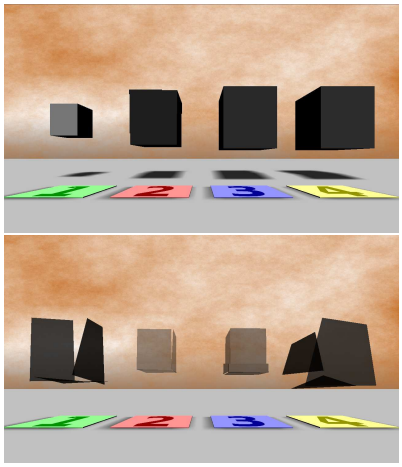

Figure 8. Third experiment: Combinations - Screenshots: two and four combined mappings

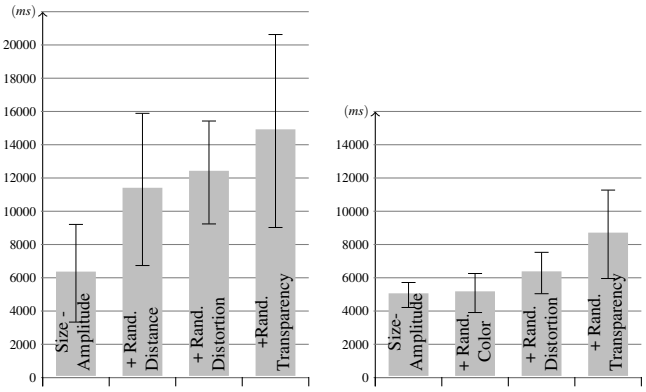

Figure 9. Third experiment: Combinations disruption - Response times (and average deviations) using Distance then Color Lightness

2.812) of the number of disrupting graphical parameters. Student-Neumann-Keuls analysis showed a significant difference between the first three sequences and the last sequence.

For the second part of this third experiment, the response times are shown in figure 10. An ANOVA on the results indicated no significant performance difference ($F = 0.368 < F(3,45,0.05) = 2.812$) when activating one, two, three or four mappings at the same time, with color as the second graphical parameter. The same lack of significance ($F = 0.550 < F(3,45,0.05) = 2.812$) was obtained with distance as the second parameter.

3.7.4. Subjects comments

Subjects reported having troubles to perceive size variations when the distance parameter was also used. They also felt

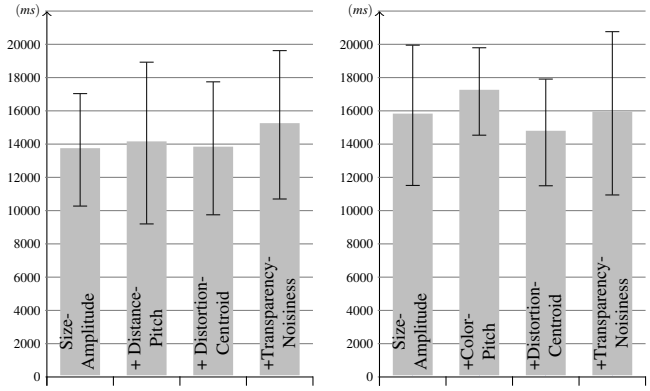

Figure 10. Third experiment: Combinations - Response times (and average deviations) using Color Lightness then Distance for the Pitch parameter

there was a problem with distance perception, because of the camera point of view. They found it more difficult to select the correct object when the variation of an audio parameter was heard but not visualised, as in the first sequence of the combination part, than when graphical parameters were randomly variating.

3.7.5. Discussion

One should notice that the response times are much longer than in the previous experiments. This may be explained by the fact that musical sequences were more complex and that it took more time to understand them. Some subjects were also disrupted by the combination of graphical parameters, to which they were not used.

The first part of the experiment revealed that the performances times dropped when a graphical parameter disrupted the perception of another. This was obvious with distance and size, and in the second test with color and transparency. This confirms the importance of the independence of perceptual dimensions, which was pointed out by Giannakis [4].

However, the second part of the experiment did not indicate any significant difference when combining up to four mappings on the objects. This part was different from the first one, because the graphical parameters variations did not always occur at the same time, and because they were connected to audio parameters variations. The results were not what we expected, i.e the combinations do not improve the performances. However, if the graphical parameters do not interfere with each other, these results suggest that may have at least four audio parameters displayed on a single object without performance loss.

4. CONCLUSION

Our results for the preferences experiment approximatively followed the results of previous user studies on musicians. Significant preferences indeed appeared for the Amplitude / Size and Spectral Centroid / Color Lightness mappings. These results also showed tendencies for the Pitch / Distance and Noisiness / Texture Rugosity mappings. The performances experiment, which was designed to measure the efficiency of the audiovisual mappings, revealed almost no significant differences, except for mappings with wrong scales. These results, combined with subjects comments, may indicate that the choice of the scales of the mappings is actually more important for the efficiency than the choice of the mappings themselves. The results of the combinations experiment did not indicate any significant drop in perfomances when using several mappings on the objects, except when some graphical parameters disrupted the perception of others. This may indicate that musicians can deal with at least four independent audiovisual mappings on a single object without performance loss, if the scales and mappings are correctly chosen. In a 3D environment, where one may navigate in large sets of audiovisual objects, this may bring new musical possibilities.

However, these results may be strongly due to the fact that subjects were musicians, used to specific visualizations. This prevents us from generalizing our conclusions to non-musicians. The results may also depend on the scales that we used.

Future studies should thus investigate which are the most efficient scales for the audiovisual mappings, taking inspiration from the information visualisation field. Other graphical and audio parameters should also be tested, for example higher level musical parameters like tempo, tonality, rythm and so on. Complex 3D shapes could also be useful to display sound parameters including several dimensions such as sound spectrum.

5. REFERENCES

[1] A. Chaudhary and A. Freed, "Visualization, Editing and Spatialization of Sound Representations using the OSE Framework," *Audio engineering society*, 1999.

[2] D. S. Ebert, R. M. Rohrer, C. D. Shaw, P. Panda, J. M. Kukla, and D. A. Roberts, "Procedural shape generation for multi-dimensional data visualization," in *Proceedings of Data Visualization 99*, 1999.

[3] K. Giannakis, "A comparative evaluation of auditory-visual mappings for sound visualisation," *Organised Sound*, vol. 11, no. 3, pp. 297–307, 2006.

[4] K. Giannakis and M. Smith, "Towards a theoretical framework for sound synthesis based on auditory-visual associations," in *University of Birmingham, UK*, 2000, pp. 87–92.

[5] ——, "Imaging soundscapes: Identifying cognitive associations between audatory and visual dimensions," in *Musical Imagery. Swets and Zeitlinger*, 2001, pp. 161–179.

[6] C. G. Healey, "Building a perceptual visualisation architecture," *Behaviour and Information Technology*, vol. 19, pp. 349–366, 2000.

[7] C. G. Healey, K. S. Booth, and J. T. Enns, "High-speed visual estimation using preattentive processing," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. 2, pp. 107–135, 1996.

[8] K. Jensen, "Timbre models of musical sounds," Ph.D. dissertation, Department of Computer Science, University of. Copenhagen, 1999.

[9] J. Krimphoff, S. McAdams, and S. Winsberg, "Caractérisation du timbre des sons complexes. ii: Analyses acoustiques et quantification psychophysique," *Journal de Physique 4*, pp. (C5):625–628, 1994.

[10] E. Lemi and A. Georgaki, "Reviewing the transformation of sound to image in new computer music software," in *Proceedings of the 4th Sound and Music Computing Conference*, 2007.

[11] S. D. Lipscomb and E. M. Kim, "Perceived match between visual parameters and auditory correlates: an experimental multimedia investigation," in *Proceedings of the 8th International Conference on Music Perception and Cognition*, 2004.

[12] S. J. Luck and E. K. Vogel, "The capacity of visual working memory for features and conjunctions," *Nature, Vol 390*, pp. 279–281, 1997.

[13] M. S. Puckette, T. Apel, and D. D. Zicarelli, "Real-time audio analysis tools for pd and msp," in *Proceedings of the International Computer Music Conference*, 1998.

[14] R. Walker, "The effects of culture, environment, age, and musical training on choices of visual metaphors for sound," *Perception and Psychophysics Vol 42(5)*, pp. 491–502, 1987.

[15] J. Whitney, *Digital Harmony: on the Complementarity of Music and Visual Art*. McGraw-Hill Inc., 1980.

[16] E. Zwicker, "Subdivision of the audible frequency range into critical bands," *The Journal of the Acoustical Society of America, vol. 33*, p. 248, 1961.