Creating shapes: musicians' and non-musicians' visual representations of sound

Mats Küssner
King's College London
Department of Music
Strand
London WC2R 2LS
+44 (0)7551 309 248
mats.kussner@kcl.ac.uk

ABSTRACT
Visualising sound and music can give rise to valuable insights into music cognition. In this experiment, musicians and non-musicians were presented with pure tones, systematically varied in pitch, loudness and tempo, as well as two short musical excerpts. Visual responses were captured using an electronic graphics tablet, and continuously acquired data of the position and pressure applied to the pen were checked for correlation with characteristics of the sound stimuli. Based on participants' verbal reports it was revealed that pitch was represented with height on the tablet, and loudness with the thickness of the line. Although both musicians and non-musicians rated themselves as fairly highly consistent in applying these strategies, and did not find the tasks too difficult, correlation analysis revealed that musicians were more accurate than non-musicians. Results are discussed in terms of perceptual and motor skills in musicians.

Keywords
Cross-modality, graphics tablet, musicianship, shape, sound.

1. INTRODUCTION
Visual aspects of music and musical performances have attracted the attention of researchers from a large number of fields [8]. In the field of music psychology, people are particularly interested in music perception and cognition, and asking participants to visualise what they hear is one way to get access to mental processes related to sound and music. While all music is sound, not all sound is necessarily music. However, when studying music empirically, it is extremely important to have a good understanding of its basic properties, especially if one attempts to tap into a new or largely under-researched area. The study of visualizations of sound and music is such an area. Although the first experiments date back at least to the 1970s [7], there have been only a few studies since then which have investigated adults' visual responses to sound and, to an even lesser degree, music. A great number of studies have been conducted with children, trying to illuminate the cognitive development of music perception. As this is not of paramount importance here, the reader may be referred to M. Barrett and J. Bamberger as a starting point [1, 2]. Walker was the first to study visual metaphors of sound in a systematic manner by testing groups of participants differing in age, cultural background, environmental conditions and musical training. What he found more or less consistently was that people tend to associate pitch with vertical space, loudness with size, timbre with pattern, and duration with horizontal space [14]. Even though his first set of experiments [13] involved free visualizations of sound by using a paper-and-pencil paradigm (the participants were children), the method of his main study was a forced-choice paradigm. To the best of my knowledge, no-one has ever systematically investigated adults' responses to basic sound characteristics in a free drawing paradigm. (Their focus being on movement rather than explicit visualizations, Godøy and Haga came closest by investigating a wide range of sounds, including many from Pierre Schaeffer's sound typology [5, 6].) Apart from investigating basic features of music by means of pure tones it is of course essential to study more musical examples in the form of musical excerpts or even whole compositions. The latter has been done by [10] who found that musically trained participants tend to represent music in an abstract manner, whereas non-musicians prefer pictorial, and potentially more diverse, representations. Only one further study [9] has attempted to gain insight into the mental processes underlying the perception of basic musical properties (in this case rhythm) by investigating musicians' visual representations. The focus of the present study being on pitch and loudness, suffice it to say that musically trained participants attended more to the underlying beat of rhythmic sequences, which became apparent in their visualizations.

The neglect of free visualization paradigms that study adults', and more specifically, musicians' and non-musicians' responses to sound and music occurs thus on many levels, and some but far from all will be dealt with here. Apart from tackling the very basic question of how musicians and non-musicians represent fundamental properties of sound visually, this experiment introduces a new methodology by studying participants' responses in real time while they listen to the sound stimulus. This means that the focus is not only on the final *product* of the visualization but also specifically on the *process* of it.

2. OBJECTIVES
Comparisons between musicians' and non-musicians' visual representations of sound by means of a graphics tablet aim to reveal any commonalities or differences that may occur as a result of training. Moreover, the present study examines which sound characteristics are most salient to listeners and how they represent them in a free, real-time visualization paradigm. The exact measurement of 'drawing performances' with the electronic graphics tablet is intended to reveal the extent to

which musicians' and non-musicians' visualizations correspond systematically to sound parameters such as pitch and loudness.

3. METHODS

3.1 Participants

The total number of participants who took part in the study was 73 (42 female), and the mean age was 28.51 years (*SD* = 7.71). The subgroup of non-musicians consisted of 31 participants (18 female) with a mean age of 28.34 years (*SD* = 7.11). There were three inclusion criteria for this subgroup: the participants must not have exceeded Grade 1 of the ABRSM examination system (http://www.abrsm.org/); they must have stopped playing more than six years ago; the overall time spent actively making music must not have exceeded two years. As for the 41 musicians (24 female; mean age: 28.63 years, *SD* = 8.23), they comprised subgroups of 12 keyboard players, 8 string players, 8 wind/brass players, 8 composers and 5 singers. Though not all of them took ABRSM or equivalent exams (there is, for example, no such grading system for composers as there is for instrumentalists), the general level among those who did was fairly high: one musician was at grade 7, 15 musicians were at grade 8, and 10 were above grade 8.

3.2 Materials

3.2.1 Sound

The sound stimuli consisted of 18 pure tones systematically varied in pitch, loudness and tempo, as well as two musical excerpts, all of which were between 5 and 15 seconds long. The pitch range from B2 (123.47 Hz) to D4 (293.67 Hz) was chosen according to the melodic line of the first bar of Chopin's prelude op. 28, no. 6. The three simplest sound stimuli consisted of single pure tones (D4) sounding for 5 seconds, either decreasing and increasing in loudness (decr-incr; SOUND_01); without change in loudness (equal); or increasing and decreasing in loudness (incr-decr). The same loudness patterns were applied to all pure tone sequences. 3 of the 15 stimuli varying in pitch showed a falling-rising contour in semitone steps (SOUND_05) whereas the 12 remaining stimuli displayed a rising-falling contour. 9 of these used semitone steps, and 3 used the notes B2, D3, F#3, B3, D4 in line with the Chopin prelude (SOUND_11). 6 of the stimuli displaying a rising-falling contour and moving in semitone steps were furthermore manipulated regarding the tempo. Half of them showed two decelerations (SOUND_14), and the other 3 two accelerations (SOUND_17). Finally, two recordings of the first two bars of Chopin's prelude op. 28 no. 6, one by Cortot from 1926 (SOUND_Cortot) and one by Argerich from 1975 (SOUND_Argerich), were included too.

3.2.2 Graphics tablet

A commercial electronic graphics tablet (Wacom^{TM} Intuos4^{TM} L) was identified as an appropriate hardware device for capturing the responses in two spatial dimensions and the pressure applied with the pen. It was also important that this tablet had a high sampling rate (maximum 200 points per second [12]) to ensure that the software did not miss a drawn point by virtue of the tablet sampling too slowly. In addition, to capture subtle positional changes, a high resolution was required in both space and pressure sensitivity (this tablet has 5,080 lines per inch and 2048 levels of pressure [12]). The capture software was developed in Processing [4] by Nicolas Gold.

3.3 Procedure

After signing the consent form, participants read a detailed instruction and were familiarised with tablet and pen. At all times, they were able to see what they were drawing on a screen in front of them (see Fig. 1). Applying more pressure to the pen resulted in a thicker line on the screen. All sound stimuli were presented with a commercially available set of headphones (Sony MDR-7506 Professional Dynamic Stereo Headphones).

Figure 1. Experimental setup: visualising sound by means of graphics tablet.

During a practice trial prior to the experimental trial, participants were presented with 9 sample stimuli very similar to those used in the proper experiment in order to get familiar with the procedure. Each sound stimulus was played twice consecutively. During the first presentation participants were asked to listen carefully without drawing, and during the second presentation they were asked to draw along with the sound as it was played, and in this way represent the sound visually. They were told that they "should draw everything in one go, that is, without going back to parts which you have already drawn. You can think of it as a musical performance, whereby it is impossible to go back in time and play the last bar or note again. Also, please try to take into account all sound characteristics you are able to identify and ideally, represent them in your drawing." Moreover, they were instructed not to use any formal or conventional symbols, numbers or letters in any of their drawings.

The order of the stimuli was pseudo-randomised by grouping the 20 experimental sound stimuli into four categories. The first group consisted of the 3 stimuli varying in loudness only, the second one contained the 9 stimuli varying in pitch and loudness (but not tempo), the third group contained the 6 stimuli varying in pitch, loudness and tempo, and the fourth group comprised the two musical excerpts. The order of the four categories was the same for each participant, whereas the order of the stimuli within each category was randomised.

Apart from recording participants' drawings the experimenter gathered information by means of a questionnaire about demographic data, musical habits (listening and/or playing), musical education, language skills, familiarity with graphic scores, and any visual or hearing impairments. Furthermore, in a short feedback interview, participants were asked whether they used any strategies in representing pitch, loudness and time (and if so, what they were and how consistently they thought they had applied them), how difficult they found the task, whether they liked or disliked anything in particular, and whether they had any other comments on the experiment. Finally, participants were thanked for their participation in the experiment and received a small amount of money for reimbursement.

3.4 Analysis

3.4.1 Selection of participants for correlation analysis

Due to participants' verbal reports regarding the strategies they used to represent pitch and loudness (see results section: pitch was most commonly associated with height, and loudness with thickness of the line), it was necessary to make a selection of participants to achieve equally sized subsamples of musicians and non-musicians for the purpose of a correlation analysis. This was carried out in four steps, beginning with a visual inspection of the drawings to find out whether participants' verbal reports were in accordance with their respective drawings. If a participant was not aware of using any strategy (or did not report it) but their drawings clearly suggested that they had applied the strategies outlined above, they were included in the analysis. Since still too many musicians were left, a stricter criterion of musicianship was used, excluding all musicians below grade 8 and with less than 2 hours of practice per week. This was sufficient for the correlation analysis between pitch and height, but not for the correlation analysis between loudness and thickness of the line, where an even stricter criterion of 'not less than 4 hours of practice per week' was necessary. Note that although the subsamples of musicians and non-musicians for the two correlation analyses largely overlapped they were not exactly the same.

3.4.2 Correlation analysis

According to participants' verbal reports two correlations were calculated: the first between pitch and height on the tablet, and the second between loudness and thickness of the line, which was proportional to the pressure applied to the pen. A non-parametric estimate (Spearman's rho) was calculated due to the serially correlated nature of the data sets. This procedure has been described in an experiment by [11], where the authors, similarly to this experiment, acquired data continuously.

The feature extraction of pitch and loudness was carried out in Praat [3]. As for the two musical excerpts, an automatic pitch extraction was run, which was later adjusted manually to make sure only the melodic line of the left hand was included in the analysis. Pitch was represented on a log-transformed frequency scale in Hertz, and perceived loudness values were measured in sone. Before being entered into a correlation analysis, all values were normalised with mean 0 and standard deviation 1. Similarly, the y values (representing height on the tablet) and pressure values (representing thickness of the line/pressure applied to the pen) recorded with the graphics tablet during the drawing experiment were rescaled to obtain 0 for the mean and 1 for the standard deviation. Both MATLAB (R2010a, The MathWorks) and SPSS (Version 19.0.0, IBM SPSS Statistics) were used for the analysis.

4. RESULTS

4.1 Verbal reports

4.1.1 Strategies for representing pitch and loudness

Asking participants after the drawing experiment whether they had used any strategies to represent pitch and loudness revealed that the majority of musicians and non-musicians used height on the tablet to represent pitch (higher on the tablet referring to higher pitches). 98% of the musicians reported using this strategy, and only one musician (2%) reported that they had used pressure to represent pitch (the lower the pitch the more pressure applied). A comparably high number of non-musicians (81%) chose to represent pitch with height, 3% used a mixed strategy of height and pressure, 3% used a completely different approach (representing feelings), and 13% reported they did not use any strategy.

Regarding the strategies for representing loudness, similar proportions were observed. 88% of the musicians reported that they used the thickness of the line (or pressure applied) to depict loudness (the louder the sound the more pressure applied) while 5% applied a mixed strategy consisting of thickness of the line and drawing circles differing in size. Two musicians (5%) chose different approaches, one of them using more pressure for softer sounds and the other using height (higher for louder sounds). One musician did not use any strategy at all. The picture looks somewhat more mixed for non-musicians, though again the majority of 56% reported that they used thickness of the line to represent loudness (thicker line for louder sounds), and 16% used thickness and something else. 22% chose to represent the loudness with a completely different strategy, and 6% did not use (or were unaware of using) any particular strategy.

4.1.2 Consistency and difficulty

It is important to report here that there were no statistically significant differences between the subsamples of musicians and non-musicians selected for the correlation analysis with regards to the perceived consistency of applying their strategies ($t(57) = .986$, $p > .30$), and with regards to the perceived difficulty of the tasks ($t(59) = .857$, $p > .30$). Both musicians and non-musicians reported that they acted fairly consistently in applying their respective strategies, and that they found the tasks relatively easy (see Fig. 2).

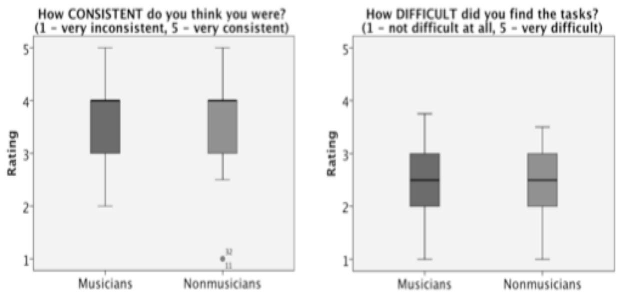

Figure 2. Box plots of participants' consistency and difficulty ratings.

4.2 Non-parametric correlations

An overview of all correlations can be seen in Table 1. Three pitch-height correlations for stimuli with constant pitch were excluded due to the *sine qua non* that variables must not be constant. The loudness-pressure correlation for SOUND_02 was omitted for the same reason. It is worth noting that all loudness-pressure correlations, whereby the amplitude of the sound stimuli remained equal but the pitch was changed, were included in the analysis since the perceived loudness changed as a function of pitch.

To find out whether musicians achieved significantly larger correlation coefficients compared to non-musicians, two independent sample t-tests were calculated for the pitch-height and loudness-pressure correlation coefficients respectively. It was revealed that musicians achieved on average a marginally significantly larger value (mean_rho = 0.851) compared to non-

Table 1. Non-parametric correlations (Spearman's rho) for pitch - height and loudness - pressure

Number	Characteristics of sound stimulus	Type of correlation	musicianship	
			musicians	non-musicians
SOUND_01	pitch: constant, amplitude: decr-incr	pitch/height	n.a.	n.a.
		loudness/pressure	.619**	.215*
SOUND_02	pitch: constant, amplitude: constant	pitch/height	n.a	n.a.
		loudness/pressure	n.a.	n.a.
SOUND_03	pitch: constant, amplitude: incr-decr	pitch/height	n.a.	n.a.
		loudness/pressure	.874**	.671**
SOUND_04	pitch: down-up, amplitude: decr-incr	pitch/height	.834**	.840**
		loudness/pressure	.634**	0.162
SOUND_05	pitch: down-up, amplitude: constant	pitch/height	.835**	.791**
		loudness/pressure	.494**	.358**
SOUND_06	pitch: down-up, amplitude: incr-decr	pitch/height	.802**	.741**
		loudness/pressure	.666**	.628**
SOUND_07	pitch: up-down, amplitude: decr-incr	pitch/height	.761**	.900**
		loudness/pressure	.344**	-0.065
SOUND_08	pitch: up-down, amplitude: constant	pitch/height	.812**	.765**
		loudness/pressure	.872**	0.047
SOUND_09	pitch: up-down, amplitude: incr-decr	pitch/height	.801**	.811**
		loudness/pressure	.951**	.572**
SOUND_10	pitch: up-down, amplitude: decr-incr	pitch/height	.769**	.702**
		loudness/pressure	0.013	-0.112
SOUND_11	pitch: up-down, amplitude: constant	pitch/height	.781**	.626**
		loudness/pressure	0.176	0.05
SOUND_12	pitch: up-down, amplitude: incr-decr	pitch/height	.702**	.520**
		loudness/pressure	.824**	.393**
SOUND_13	pitch: up-down, amplitude: decr-incr, tempo: dec-dec	pitch/height	.894**	.892**
		loudness/pressure	.733**	.532**
SOUND_14	pitch: up-down, amplitude: constant, tempo: dec-dec	pitch/height	.938**	.955**
		loudness/pressure	.271**	.557**
SOUND_15	pitch: up-down, amplitude: incr-decr, tempo: dec-dec	pitch/height	.895**	.923**
		loudness/pressure	.980**	.901**
SOUND_16	pitch: up-down, amplitude: decr-incr, tempo: acc-acc	pitch/height	.966**	.915**
		loudness/pressure	.790**	.701**
SOUND_17	pitch: up-down, amplitude: constant, tempo: acc-acc	pitch/height	.986**	.869**
		loudness/pressure	.810**	.328**
SOUND_18	pitch: up-down, amplitude: incr-decr, tempo: acc-acc	pitch/height	.987**	.876**
		loudness/pressure	.973**	.952**
SOUND_19	Chopin's prelude No. 6 (Argerich, 1975)	pitch/height	.800**	.382**
		loudness/pressure	.360**	-0.11
SOUND_20	Chopin's prelude No. 6 (Cortot, 1926)	pitch/height	.912**	.695**
		loudness/pressure	.211**	-0.07

n.a. No correlation possible because at least one variable remains constant.
*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
decr = decrease, incr = increase, dec = deceleration, acc = acceleration

musicians (mean_rho = 0.777) for the pitch-height representations ($t(32) = 1.747$, $p = 0.090$, two-tailed). Regarding the loudness-pressure correlations, musicians (mean_rho = 0.610) showed significantly larger values on average compared to non-musicians (mean_rho = 0.353) at the 0.05 significance level ($t(36) = 2.461$, $p = 0.019$).

Figure 3 offers an overview of 36 randomly chosen drawings (out of 61 in total) in response to SOUND_08. The drawings' pitch-height correlation coefficients increase from top to bottom and left to right, with most of the musicians' drawings (indicated by the rectangles) agglomerating in the lower part.

5. DISCUSSION

When asked to represent pure tones varying in pitch, loudness and tempo, as well as two short excerpts from a Chopin prelude, by means of a graphics tablet, the majority of musicians and non-musicians reported that they used the height on the tablet to represent pitch (higher on the tablet referring to higher pitches), and the thickness of the line achieved by the pressure applied to the pen to represent loudness (more pressure for louder sounds). Overall, both groups showed relatively high consistency and relatively low difficulty ratings, indicating that musicians and non-musicians thought that they applied their respective representation strategies consistently throughout the experiment, and that the tasks were quite easy. In light of these findings, it was revealed, however, that musicians outperformed

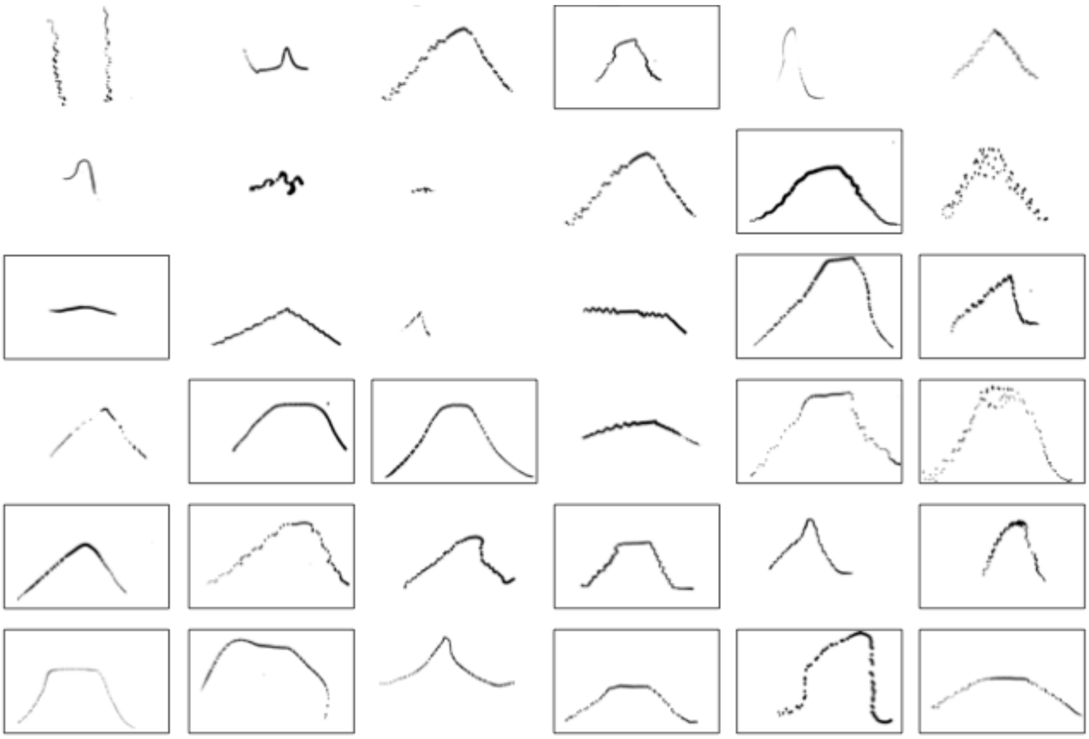

Figure 3. 36 (out of 61) randomly chosen visual representations of SOUND_08. The drawings' pitch-height correlation coefficients increase from top to bottom and left to right. Rectangles indicate musicians' drawings.

non-musicians regarding the accuracy of their representations. Non-parametric correlations between the sound and drawing characteristics showed that musicians are overall more accurate in representing loudness with the thickness of the line than non-musicians. A similar picture was obtained for the representation of pitch with the height on the tablet. Although a two-tailed *t*-test strictly speaking failed the significance level, deploying a directed hypothesis (musicians are better than non-musicians in representing sound characteristics visually) reaches significance at the 0.05 level. It is important to emphasise that a strong correlation depends on two factors. The first is related to the *products* of the visualization, many of which one is able to examine in Fig. 3. But even the most accurate representation of either pitch or loudness does not achieve a large correlation coefficient if the *process* of the visualization comes too early or too late. In other words, the timing of the drawing plays a pivotal role here, as does the ability to coordinate the anticipation of the progress of the sound stimuli with the motoric output.

Musicians' superiority became, perhaps not surprisingly, particularly obvious in the short Chopin excerpts. One could argue that this is due to their familiarity with, and hence their ability to anticipate, these pieces, but it should be noted that all participants listened in the practice and experimental trials at least three times to these excerpts before they were asked to draw along.

The fact that musicians chose to represent pitch and loudness the way they did, and that non-musicians showed greater variability in their responses is not extremely surprising and has been reported elsewhere [14]. However, the fact that subgroups of musicians and non-musicians who apply the same strategies, evidently consistently and confidently, differ in part largely with regards to the accuracy of representing very simple (non-musical) sound stimuli is somewhat unexpected. Three explanations for the observed phenomenon seem plausible. These have either to do with a perceptual hearing ability; the motor control of the pen; or a combination of both, which is the most likely candidate. Given that most of the sound stimuli were simple, monophonic sequences of pure tones it seems likely though that perceiving the sounds correctly was less of an issue. On the other hand, the influence of musicians' non-specific motor skills coupled with anticipatory listening skills might well explain the observed variance. Future experiments should compare groups of participants with varying degrees of musical and drawing skills, such as visual artists (with and without musical training), singers (without highly trained motor skills of their hands), and any group of instrumentalists whose members need extensive fine-tuned motor skills to master their instrument.

6. CONCLUSION

One should caution against drawing premature conclusions in light of the current data set. This study tried to establish a new experimental paradigm in testing musicians' and non-musicians' free visualizations of sound with an electronic graphics tablet, thereby going beyond previously applied forced-choice and paper-and-pencil paradigms. In an attempt to come up with a new methodology of capturing visual responses to sound, it was revealed that differences between musicians and non-musicians are sometimes hidden when studying the final *product* of the visualization. It is suggested that investigating the *process* of sound visualizations can give valuable insights into covert aspects of music cognition.

7. ACKNOWLEDGMENTS

I am grateful to the AHRC and CMPCP for financial support that enabled my research to take place. I also owe a huge debt to the whole 'Shapes' team at KCL, Daniel Leech-Wilkinson, Helen Prior, Dan Tidhar and Nicolas Gold (UCL), whose continuous support and constructive feedback is immensely valuable to me. Special thanks go to John Downing (aka MMB star) for correcting a first draft of this paper.

8. REFERENCES

[1] Bamberger, J. *The mind behind the musical ear: How children develop musical intelligence*. Harvard University Press, Cambridge, MA, 1995.

[2] Barrett, M. Representation, cognition, and communication: Invented notation in children's musical communication. In *Musical communication*, Miell, D., MacDonald, R., and Hargreaves, D., Eds., Oxford University Press, Oxford, 2005.

[3] Boersma, P., and Weenink, D., "Praat: doing phonetics by computer," Version 5.2.39, 2011.

[4] Fry, B., and Reas, C. (2011, 15th July 2011). *Processing*. Available at: http://processing.org

[5] Godøy, R. I., Haga, E., and Jensenius, A. R. Exploring music-related gestures by sound-tracing. A preliminary study. Paper presented at *2nd ConGAS International Symposium on Gesture Interfaces for Multimedia Systems*, University of Leeds, 2006.

[6] Haga, E. *Correspondences between music and body movement*. Ph.D. thesis in Musicology, University of Oslo, Oslo, 2008.

[7] Hooper, P. P., and Powell, E. R. Influences of musical variables on pictorial connotations. *Journal of Psychology, 76*, 1 (1970), 125-128.

[8] Küssner, M. B. Shaping music visually. In *Kreativität - Struktur und Emotion* (Würzburg, 2010). Königshausen & Neumann, Würzburg, forthcoming.

[9] Smith, K. C., Cuddy, L. L., and Upitis, R. Figural and metric understanding of rhythm. *Psychology of Music, 22*, 2 (1994), 117-135.

[10] Tan, S.-L., and Kelly, M. E. Graphic representations of short musical compositions. *Psychology of Music, 32*, 2 (2004), 191-212.

[11] Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. Cross-modal interactions in the perception of musical performance. *Cognition, 101*, 1 (2006), 80-113.

[12] Wacom™. (2011, 18th July 2011). *Intuos4 Series Product Brochure*. Available at: http://www.wacom.eu/_bib_user/dealer/bro_int4_en.pdf

[13] Walker, R. Perception and music notation. *Psychology of Music, 6*, 1 (1978), 21-46.

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

9. APPENDIX

Musical recordings used in the experiment:

1. SOUND_Cortot: Alfred Cortot: Chopin, Prelude in B-minor, Op. 28 no. 6. HMV matrix Cc-8157-3 (recorded 23 March 1926), issued on HMV DB 957. Transfer © King's College London 2007.

2. SOUND_Argerich: Martha Argerich: Chopin, Prelude in B-minor, Op. 28 no. 6 (recorded October 1975). Deutsche Grammophone 1977.