Sound and shape: implementing findings from crossmodal research.

Matei Szabo

Abstract— **Cross-modal correspondence research typically studies the association between shape and sound in separate experiments. This research aimed to explore these associations when they are integrated into a single form. Characteristics like acuteness, brightness and size of shapes were used as variables for creating an animated shape in an interactive questionnaire. Participants were asked to choose one shape out of three which they found most appropriate for the sound they heard. Analysis of the data shows that certain variables were weakly but significantly correlated to sound features, while others could not be predicted. These results provide a framework for further research both from a scientific and artistic perspective.**

1. INTRODUCTION

Music visualization attempts to represent music by mapping metrics like spectral information and loudness to some visual representation such as color and shape. These visual characteristics are defined by the programmer or artist according to their own preferences. There are however no approaches that use knowledge from research on associations between sound and shape. These findings would allow for the generation of representations of music that relate more directly to peoples perceptions of music that might have not yet been explored in other music visualization solutions. The question is how to implement these findings into a visualization and how to map these music. This paper presents a possible framework for approaching this challenge. First it describes the research on which this approach is based, then how it was integrated into a single design. Afterwards the musical features that were used to a extract meaningful information from music are described and a method for mapping these features on to visual characteristics is presented.

1-1 Cross-modal correspondence research

The notion that people match certain sounds with certain shapes has been around since the late 1920s. Sapir (1929) found that there is a contrast between the associations people have between made up names containing the vowels a and i, such as mal and mil. When asked to give these names to two objects which differed in size, people consistently assigned the names with a vowels to the larger values and the ones containing the i vowel with the smaller object.

One of the most well known cross-modal correspondences (CMC) the Bouba/Kiki effect first observed by Wolfgang Khler in 1929 and reproduced by Ramachandran and Hubbard (2001). The study asked both American and Indian students to match the nonsense names Bouba and Kiki to either a rounded shape or an angular shape (right). In 95-98% of the cases participants assigned Kiki to the angular shape and Bouba to the rounded one. This suggested that people associate information from the two modalities of speech and vision in a consistent way.

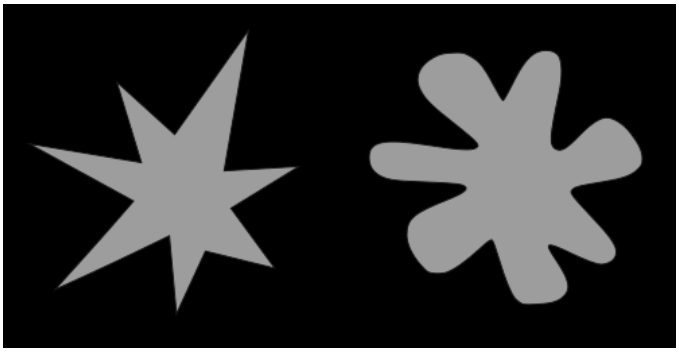

Fig. 1. Shapes used in the Ramachandran and Hubbard (2001) study

The effects of CMCs on human information processing can be seen during a speeded classification task. In these kinds of tasks participants are asked to indicate whether a presented stimulus has one characteristic or another (e.g. is this shape angular or rounded) using a method that allows for quick responses (e.g. two buttons). When two stimuli in different modalities are presented, like a shape and a sound, participants will take slightly but consistently longer to respond if the stimuli are incongruent. For example if a high pitched sound is presented together with a rounded shape, participants will take longer to classify the characteristic of the shape than if the stimuli are congruent, e.g. a high pitch and an angular shape. Figure 2 shows the experimental setup from Gallace and Spence (2006) in which participants would be presented with a fixation point for 300 milliseconds, an empty screen and the first visual stimulus without sound for 300 ms. After 500 ms, the same or a different visual stimulus would be presented for 80 ms. The task is to indicate whether the two visual stimuli (called disks in this example) were the same size or different sizes. The second time the visual stimulus was presented, a congruent, incongruent or no sound stimulus was also presented. The Reaction Times (RT) for the participants are shown to the right of the setup. The researchers measured the lowest RT for the congruent pairs of stimuli, indicating that this facilitates the performance on the task.

From these kinds of researches four correspondences have been chosen that are most suited to be integrated in a single entity:

- Angularity: sharp shapes correspond to high pitched

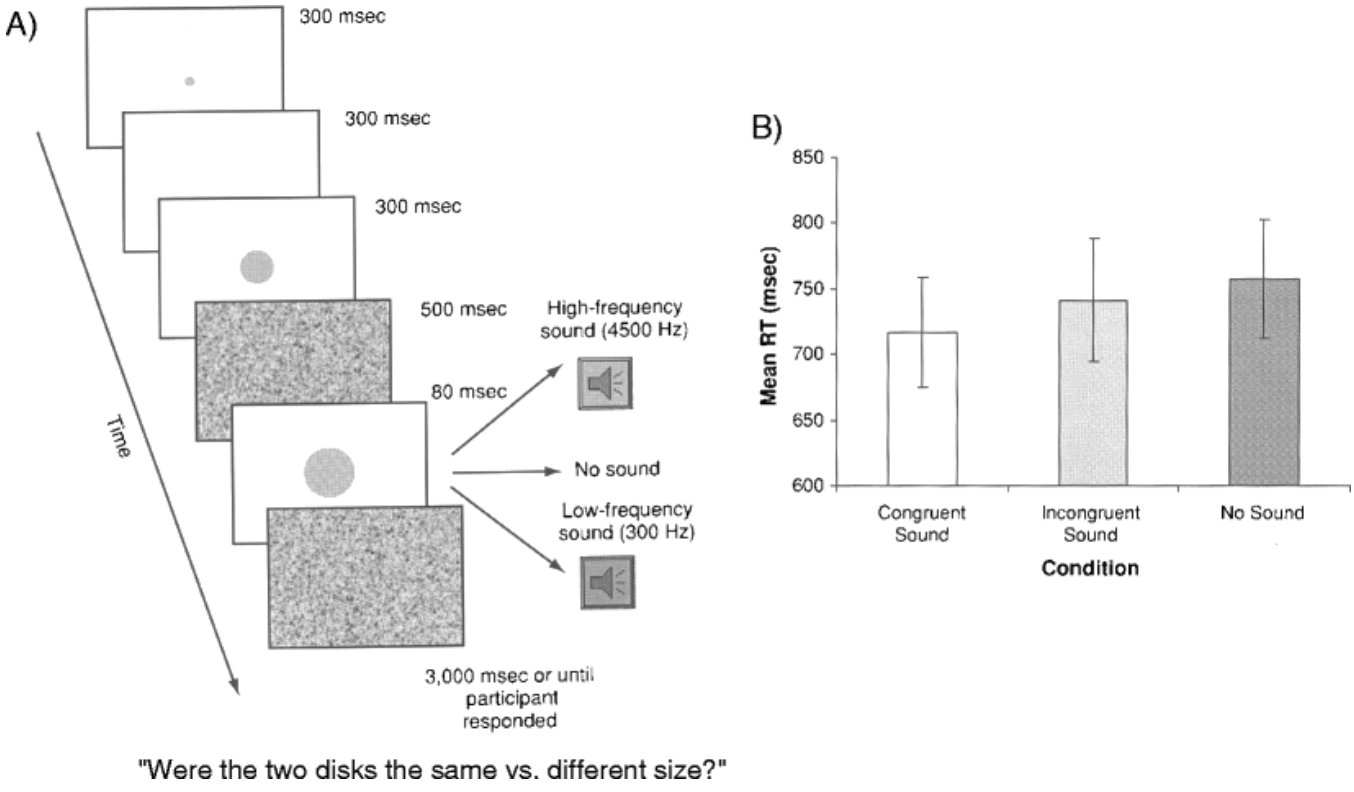

Fig. 2. Common setup for a speeded classification task

sounds and rounded shapes to low pitched sounds (Marks, 1987)

- Brightness: bright lights correspond to high pitched sounds and dim lights correspond with low pitched sounds (Marks, 1987).
- Position: high positions of stimuli correspond with high pitched sounds and low position with low pitched sounds (Evans & Treisman, 2010).
- Size: smaller stimuli correspond to higher pitches and larger stimuli to lower pitches (Evans & Treisman, 2010).

1-2 Spectromorphology

CMCs are studied in the field of psychology but in musicology attempts have been made to transcribe sounds that vary in qualities that are more complex than classical instruments. Compositions for a piano or violin can be represented as a set of pitches over time with some additional information about duration and strength. Electronically generated sounds used in compositions vary in characteristics like timbre and can be comprised of sound samples that are difficult to represent in a direct way. The field of spectromorphology attempts to create a system for expressing these characteristics visually (Paton, 2007). While spectromorphology is mostly preoccupied with the representation of live compositions of electroacoustic music, the practice of music visualization attempts to represent recorded music by mapping metrics like spectral information and loudness to some visual representation. Today visualization software is easily obtainable and often incorporated into software music players.

1-3 Current research

This research aims to investigate to what degree the findings from CMC research are replicable when the shape characteristics are integrated into a single shape and the sound stimulus is a music clip. The challenge here is that music is comprised of many sound characteristics that change over time, unlike the controlled stimuli used in CMC research. To explore these relations effective music features will be sought and extracted. These features must represent characteristics that are related to the CMC research mentioned above. This research also explores other sound features that might might help elucidate the relation between auditory and visual perception.The following hypothesis is established regarding the relation between the shape characteristics and the music features:

Higher average music clip pitch is associated with more angular shapes, brighter shapes and smaller shapes.

Position was not incorporated into the design because in CMC research the stimuli are presented one at a time with a clear fixation point in the middle, while in this research three are presented next to each other. Rather, the design of the shapes focus more on motion, an attribute that has, to our knowledge, not been studied in CMC research. For this motion, relevant sound features will be sought since it might

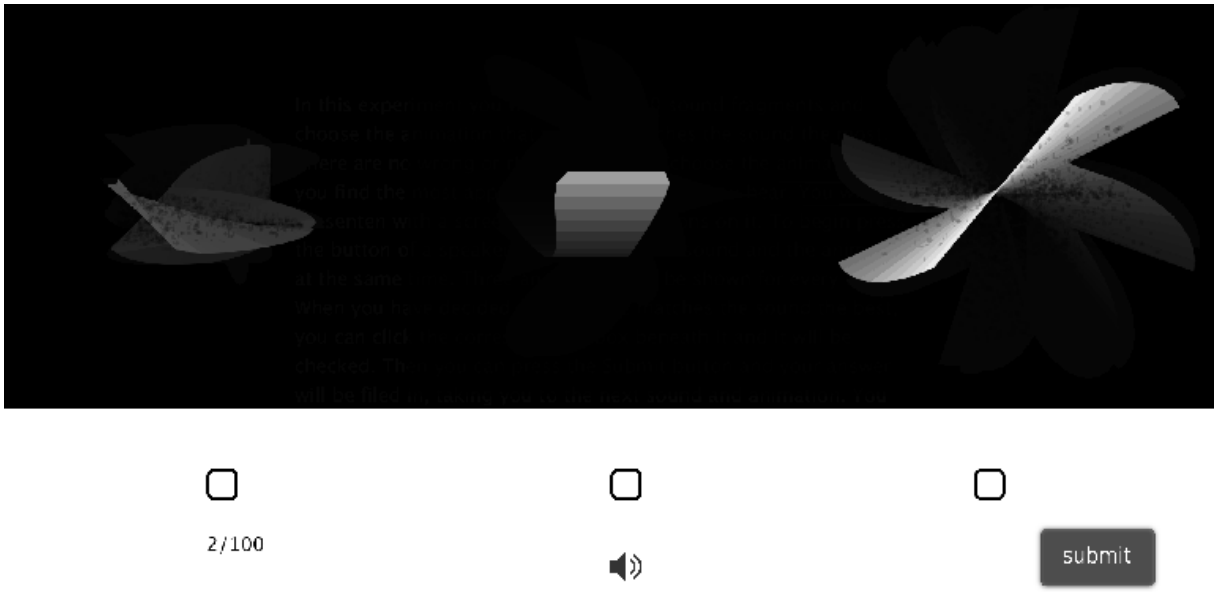

Fig. 3. Example of a questionnaire trial

not be related to pitch..

2. Methods

Processing 3 was used to generate the animated shapes, play the sound clips and export the data to .txt files of analysis. The Processing sketch runs at a resolution of 1920x1080 and a refresh rate of 60 Hz and headphones were used for the sound. It starts with instructions on how to take the test, including that there are no right or wrong answers to encourage participants to use intuition rather than systematic judgment. Afterwards participants are shown the first question asking them to choose an appropriate shape for the sound clip that they hear. They are presented with three animated shapes and a single sound clip and have the option to repeat the animation and sound simultaneously as many times as they need to make a decision. When they are sure of their choice they check the box beneath the chosen shape and press the submit button (Fig 3). The combinations of variables used to create the shapes in the first 36 (practice) trials were recorded and presented again (repeat trials) after the 64th trial, in order to check the consistency of participants judgments. After completing all 100 shape sound matches the data is written to a .txt file and they are instructed to announce that they are done.

The audio clips used in this study were obtained from a purchased copy of Enas Binaural album. This was chosen for the albums use of textures rather than pitched instruments which is more representative of the kind of electroacoustic music researched in spectromorphology. This unconventionality would also lessen participants habitual interpretations that they have with tonal music. Four songs were used in this project, with sixteen one second long samples being taken from each one. Because rhythm was beyond the scope of this research the sound clips did not span a full bar of music but were clipped to an arbitrary length of one second. This was long enough for participants to hear enough sound to make an informed decision but short enough for them not to linger on it. The sound clips were taken from the beginning of the song at intervals of 16 measures. The total amount of 100 sound clips were split into two data sets of 64 clips each: one containing the first 36 clips (referred to as ”practice trials”) and one containing the last 36 (referred to as ”repeat trials”, as they were repeats of the shapes and sound clips from the practice trials). The 28 trials in the middle were present in both data sets. The feature of sound most often used in CMC research is the pitch of a sound. In those cases the sounds are created and controlled by the experimenters. In the case of this research, some useful pitch information had to be extracted along with other features that might be useful in representing the qualities of the sound. The tool used for the feature extraction was the LibROSA 0.6 library for Python.

2-1 Inspiration for shapes

To incorporate the characteristics of shapes that are associated with sound such as angularity, brightness and motion, a flexible and easily modifiable solution was sought that would work within the performance boundaries of the Processing 3 program. For this reason, the shapes presented to participants are inspired by Lissajous curves made with oscilloscopes. The curves result from plotting the values of two periodic oscillations, one on the X axis and one on the Y axis. If the oscillations are quick enough the curves drawn on the oscilloscope leave a trace on the phosphor coated inside of the screen giving the impression of a solid shape. The interaction between the separate curves in time and the emergent and unexpected shapes created this way inspired

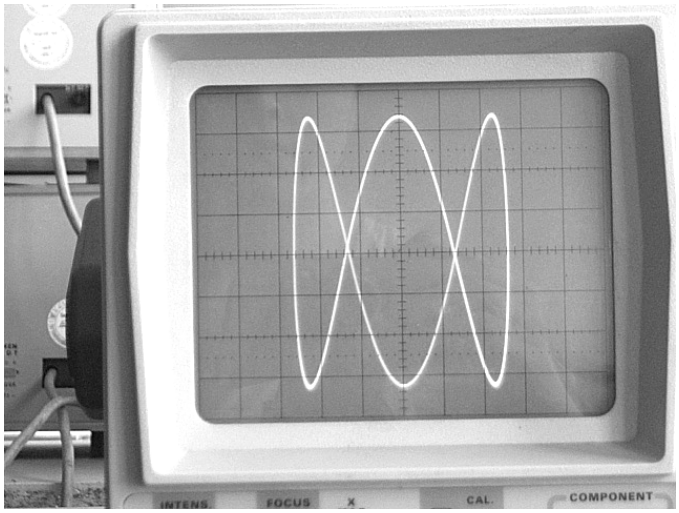

Fig. 4. Example of a Lissajous figure displayed on an oscilloscope

the design and variables that were used in this project. They were not derived from previous research, but were designed to mimic the oscilloscopes image look by rotating a elongated polygon around its center and varying its length, position on the screen and amount of color fill.

2-2 Shape variables

Angularity (spikiness) is achieved by changing the length of the polygon according to the value of a sine function. Looping through the first half of the sine wave values creates a round convex shape and a round concave spiky shape with the second half. The intermediate shape is made by halving the time of the loop. Brightness is determined by the fill() function of Processing, higher values make the shape more white. Movement is modulated by transposing the coordinates of the polygon in a repetitive manner using sine waves. Other variables include the frequency at which the polygon moves vertically or horizontally and the amount of trail left by the polygon during movement (this is achieved by placing a black rectangle over the entire polygon and varying the alpha channel of its fill). It is important to note that the motion of the shapes is not synchronized to the music. In total, 9 variables were used in the generation of the shapes and in the subsequent analysis.

When interpreting values during analysis, the following should be taken into consideration:

- Angularity: The higher the value, the more angular the shape is.
- Brightness: Higher value, brighter shape.
- Horizontal and vertical motion: higher value, greater motion.
- Horizontal and vertical speed: higher value, higher speed.
- (Counter)clockwise motion: a value of -1 result in a counterclockwise rotation, a value of 0 in no rotation and a value of 1 in a clockwise rotation.
- Fade: this value indicates the amount by which the brightness of the shape fades to black during each frame that it is drawn, the higher the value, the quicker it fades and less trail it leaves.

2-3 Motion

As mentioned, position is not utilized in the visualization but motion is integrated into it. This motion is not synchronized to the tempo of the music, rather three possible values are used to vary the speed of the motion and three others to vary the size of the screen over which the motion takes place. It was chosen to arbitrarily pick these values and randomly distribute them over every sound clip presented to participants and to see during analysis whether the related to the any of the sound features.

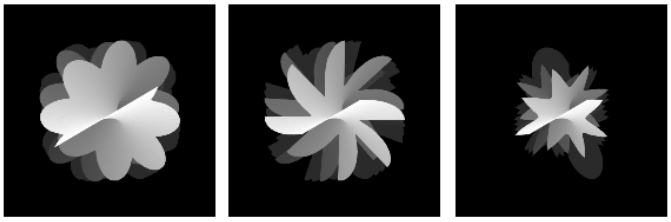

Fig. 5. Three shapes drawn with three different angularity values

Fig. 6. Three shapes drawn with different angularity and brightness values

Fig. 7. Three shapes with three different horizontal motion size values

2-4 Combinations of values per variable

To define the shapes that are presented to the participants certain parameters had to be chosen for the variables. Figure 5 shows the three different values for the variable angularity ranging from rounded to angular with all other variables being equal. Note that the rightmost shape is somewhat smaller than the two to its left. This unintended effect will be discussed in the discussion (section 4).

Figure 7 shows three shapes with varying horizontal motion sizes, from low to high. These motion sizes influence to a high degree the resulting total shapes. Figure 6 shows shapes with the same values for all variables except angularity (low

to high) and brightness values ranging from the minimum to the maximum.

For every variable three values have been chosen that intuitively differentiate the shapes enough from one another. Before a new set of three shapes are presented, the three values are randomly sampled without replacement for each variable, meaning that no value will be used twice in a single trial. This was done in order to explore as many combinations as possible for the shapes. As mentioned in 2.2, the motion of the shapes is not synced to the music, there are 3 speed and 3 movement size values that are randomly distributed per trial.

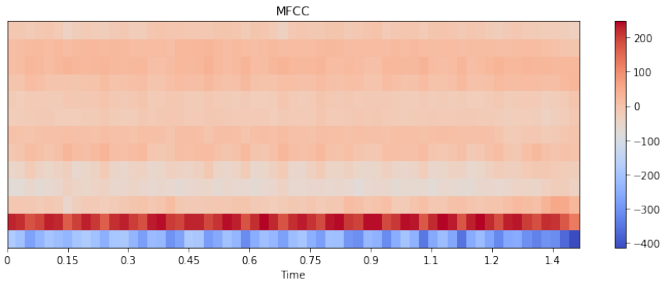

Fig. 8. A time series of MFCC's

2-5 Sound feature analysis

The simplest approach to comparing sound clips based on their frequency content is taking the average of all pitches beyond a certain frequency cutoff point. The issue is that there is no indication of where that cut off point should be in order to legitimately separate low pitches form high ones. An attempt was made with an arbitrary cutoff point at 1000 Hz using the obspy.signal.filter.highpass (Fig 9, 10).

Another solution is to use the average of the most prevalent frequency (pitch) in the analysed section of the sound, and is related to the perception of brightness of a sound (Grey & Gordon, 1978). Taking the average of all the sections in the sound clip would therefore offer a dominant frequency for the entire clip.

Spectral flatness is a measure that indicates to what extent the power of a frequency spectrum is equally distributed over all the frequency bands versus concentrated around a few bands. This measure is used to indicate the amount of noise versus tones (Dubnov, 2004). Considering the the average of spectral flatness scores was taken over the entire clip, this offers an amount to which the entire clip resembles noise.

The Mel Frequency Cepstrum Coefficient (MFCC) is a audio feature representation that is used in music information retrieval and speech recognition. The computation of the MFCC is quite complicated but its important characteristic is that it represents sound features in a way that is close to how humans process audio information. This feature is usually used in combination in a machine learning approach to sound analysis and this research has attempted to use it as well.

Harmonic salience is used to detect pitch (defined here as fundamental frequency) as an alternative to the spectral

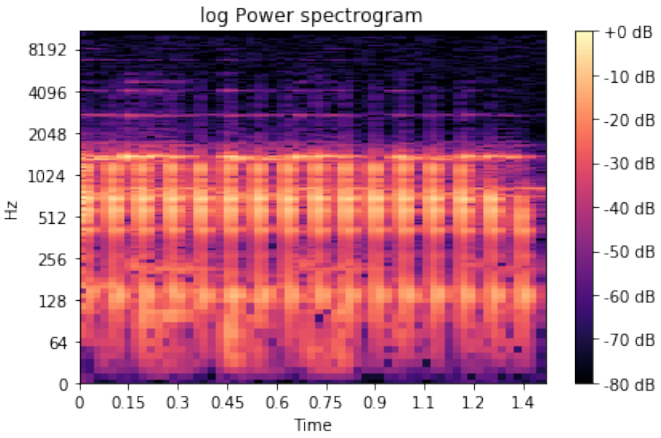

Fig. 9. Normal log power spectrogram

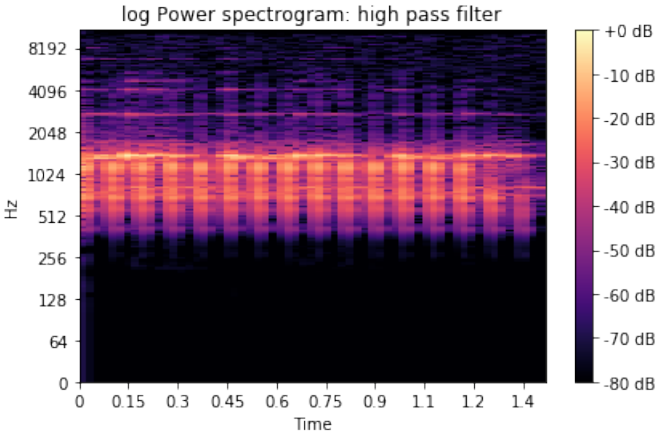

Fig. 10. Log power spectrogram of the same sound clip using a high pass filter (cutoff 1000 Hz)

centroid approach. This method results in an approximation of pitch that does not rely on frequency magnitudes and thus is less sensitive to timbre differences than the spectral centroid (Degani, Leonardi, Migliorati & Peeters, 2014) Onset detection indicates at what frames in the sound clip a new note or audio event begins based on sudden changes in the signal. For the autocorrelation analysis, which represents repetitive patterns in the sound signal and when graphed looks like a oscillating signal itself, the resulting set of numbers were analysed using a periodogram. Taking the weighted mean of these numbers indicates the dominant periodicity of the autocorrelation. Depending on the analyzed sound signal, the resulting numbers indicate the repetitiveness of percussive or harmonic sounds.

Median-filtering harmonic percussive source separation was applied with the rationale that the harmonic (tonal) part of the soundclip might be related to the variables angularity, brightness and size like in CMC research. The percussive part of the signal (comparable to tempo) was expected to relate to motion of the shape.

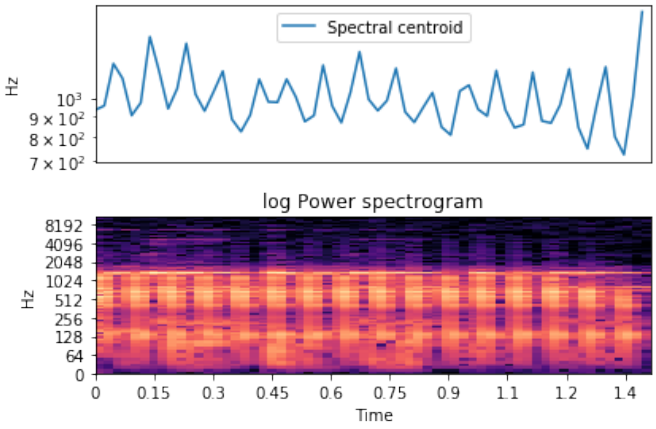

Fig. 11. Above: a time series of spectral centroids, below: a log power spectrogram of the same sound clip

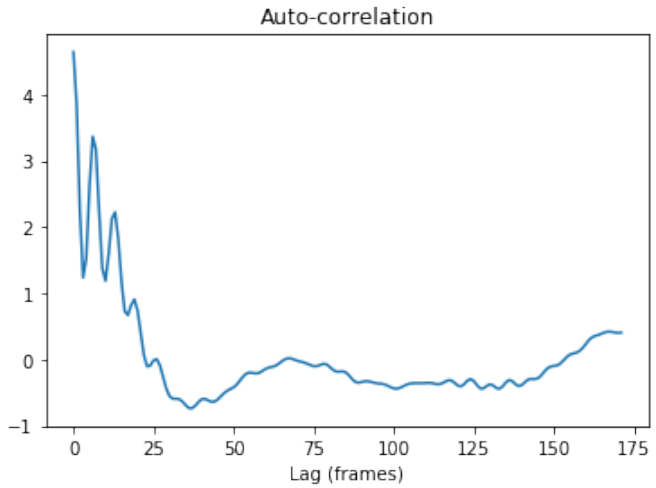

Fig. 12. A graph indicating the magnitudes of correlation by the amount of lag in the compared sound signals

3. Results

3-1 Investigating the task data

The sample consisted of mostly students recruited through the MediaTech program or through acquaintances. A total of 25 participants completed the task. Analyses were conducted using IBM SPSS version 25. The reasoning behind this separation is that participants would need some time to get used to the task (during the practice trials) and that they might form preferences the more they practiced it. The frequency distributions of the values chosen by participants for the 9 shape variables were analysed as well as their correlations with the sound features extracted from the sound clips. Additionally, correlations between shape variables were also checked.

Shapes with no rotation (coded as a value of 0) were removed from the analysis because of their drastic modulation of the entire shape, possibly masking the other variables representation.

Figure 13 shows two shapes with the same values for all variables except the rotation variable, which is 0 in the second image. Removing this variable from analyses resulted in more consistent findings, supporting that this metric negated the effect of other variables.

Observation of the frequency distributions suggest that in general participants chose certain values more often than others. To find the direction of the asymmetrical distribution, skewness was assessed for each variable. The low amount of skewness (within -.5 and .5) only shows small asymmetry. The interpretation of this test is somewhat counterintuitive: negative skewness means that the mass of the distribution curve is concentrated more to the right of the figure and vice versa for positive values. The values that show negative skewness suggest that participants chose in general: more vertical movement (-.123), faster moving vertical shapes (-.120), more horizontal movement (-.0.98) and brighter shapes (-.073). The positive skewness values suggest that participants in general chose more rounded shapes (.30), quicker fading shapes and more often counter-clockwise motion (.111). Analyses between variables show negative correlations between vertical motion and horizontal motion ($r(1065) = -.063$, $p < .05$), angularity and vertical motion size($r(1065 = -.064$, $p < .05$) and brightness with vertical motion speed ($r(1065) = -.070$, $p < .05$). This suggests that participants chose either vertical moving shapes or horizontal moving shapes, but not both. Large vertical movement was less often chosen together with angular shapes and more rapid vertical movement less often with brighter shapes. Analyses of the practice and repeat trials, which were identical in shapes and sounds, offer insight into consistency in participants choices. Larger skewness scores on the practice trials suggest that participants had more pronounced preferences for horizontal motion speed (skewness difference of .25) horizontal motion size (.20), clockwise/counterclockwise rotation (.31) and to a lesser degree vertical motion size (.15), fade (.10), and angularity (.10). Brightness (.01) and size (.01) seem to have remained consistent between the two trial blocks. Examination of the correlations between shape variables in the two blocks reveal that the only correlation from the practice trials between vertical and horizontal motion size $r(544) = 089$, $p < .05$ is flipped for the repeat trials $r(609) = -0.091$, $p < .05$, suggesting that participants were more likely to choose more

Fig. 13. Difference between two shapes with clockwise movement (left) and with a (counter) clockwise movement of 0 (left)

Table 1.

Correlations between the sound feature and shape variables

	Avg. magnitudes	Average spectral centroid		Spectral flatness	Harmonic salience		Onset detection	APWA		
		Overall	Percussive		Unweighted	Weighted		Overall	Percussive	Harmonic
Angularity		.079*				.077*			.063*	
Brightness		.065*	.083*	.064*		.071*				
Size							.062*			.064*
Horizontal motion size	-.065*	.087**	.088**		.081*			-.083*		
Vertical motion size	.064*									
Horizontal motion speed									-.08*	-.083*
Vertical motion speed										

*$p < .05$, **$p < .005$

Table 2.

Multiple regression results for the sound features per shape variable

	Spectral centroid avg.	Harmonic spectral centroid average	MFCC weighted means	Weighted harm. Salience	Harmonic salience	Harm. Autocorr	Onset count	Average highpass magnitude
Angularity		1	1					
Brightness				2				
Size						3	3	
Horizontal Motion size	4					4		4
Horizontal Motion speed						5		
Vertical motion size					6			6

Note: 1: $F(2,1064) = 5.390$, $p = .005$, $R^2 = .01$
2: $F(2,1064) = 5.431$, $p < .05$, $R^2 = .005$
3: $F(2,1064) = 4.611$, $p < .05$, $R^2 = .009$
4: $F(2,1065) = 7.360$, $p < .05$, $R^2 = .007$
5: $F(3,1063) = 5.463$; $p < .005$, $R^2 = .015$.
6: $F(2,1064) = 4.419$, $p < .05$, $R^2 = .008$

vertical and horizontal motion together in the practice trials but less in the repeat trials.

3-2 Data processing for SPSS analysis

Because the participants choices are recorded per audio clip, the audio features must also represent the sound qualities of the entire sound clip. This means that there must be a single data point per clip in order to be statistically analyzed in SPSS. The feature extraction in libROSA however, returns time series data. It chops sound clips up into smaller windows (usually a few milliseconds long) and analyzes them in order from the beginning to end of the audio. In order to extract a single data point for the entire clip, the average of all the values for all the time windows were taken per clip for some analyses. The spectral centroid for example, is returned as a series of numbers that represent the dominant frequency in an analysed window. Taking the average of all these frequencies will offer an average dominant frequency for the entire sound clip and would therefore allow sound clips to be compared on that basis (e.g. the average dominant frequency for sound clip 10 is 722 Hz and 915 Hz for sound clip 4, therefore clip 4 is higher in pitch than clip 10). Other values such as the numeric output from a mel frequency cepstrum coefficient (MFCC) is returned as magnitudes per coefficient bin at every window in the sound clip (13 coefficient bins x 61 windows = 793 data points). To represent these data as a single number, the coefficient bins were weighted by their corresponding magnitudes for all windows, yielding a weighted average of the MFCC. For the autocorrelation analysis, that represents repetitive patterns in the sound signal and when graphed looks like a oscillating signal itself, the resulting set of numbers were analysed using a periodogram. Taking the weighted mean of these numbers indicates the dominant periodicity of the autocorrelation. Depending on the analyzed sound signal, the resulting numbers indicate the repetitiveness of percussive sounds or harmonic sounds. Median-filtering harmonic percussive source separation was applied with the rationale that the harmonic (tonal) part of the soundclip might be related to the variables angularity, brightness and size like in CMC research. The percussive part of the signal (comparable to tempo) was expected to relate to motion of the shape. As the task of finding one number to represent an entire sound clip is quite reductive, we attempted several feature extractions to see if there were correlations between these extractions and any visual features of the shapes. The following section is an explanation of each method which was used which yielded significant correlations, as well as an explanation of the method.

3-3 Sound feature analysis correlations

Averaged spectral centroids correlate with horizontal motion size ($r(1065) = .087$, $p = .005$), angularity ($r(1065) =$

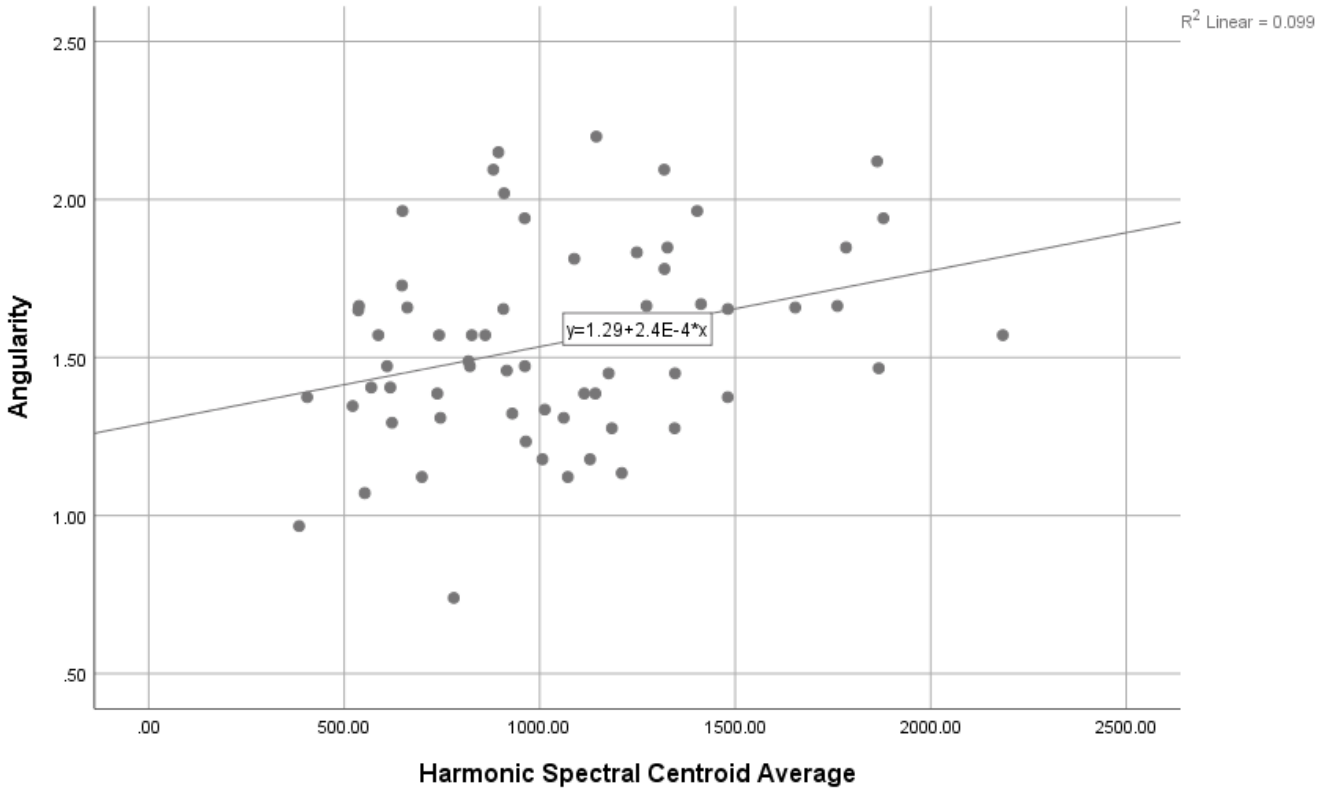

Fig. 14. Plot of angularity and spectral centroid values

.079, $p < .05$) and brightness ($r(1065) = .065$, $p < 0.05$). Spectral centroid means for the percussive components of the sound clips correlate with horizontal motion size ($r(1065) = .088$, $p < .005$) and with brightness ($r(1065) = .083$, $p < .05$). This feature was expected to relate to the angularity and brightness characteristics, as it is an indicator of average pitch height. Its relation to horizontal motion size is an unexpected result similar to the averaged highpass magnitudes. Spectral flatness this feature correlates with brightness ($r(1065) = .064$, $p < .05$), suggesting that more noise like sounds are associated with brighter shapes. MFCC An analysis was conducted with the open source machine learning suite using the sound features, including the MFCC however the analyses could not yield useful models for describing the data. Instead, the Weighted average of the MFC coefficients were used in the multiple regression analyses (discussed later). This feature did not correlate with any shape characteristics on its own. Harmonic salience correlates with horizontal motion size ($r(1065) = .081$, $p < .05$). Weighted mean harmonic salience also correlates with angularity ($r(1065) = .077$, $p < .05$) and brightness ($r(1065) = .071$, $p < .05$). These findings are in line with the spectral centroid findings. This might be due to the fact that both features are indicators of average pitch. The amount of detected onsets was summed for every clip and correlated with size ($r(1065) = .062$, $p < .05$). This is an unexpected finding considering that size is expected to relate to pitch and not temporal/rhythmic attributes. Autocorrelation periodogram weighted averages (APWA) correlated with with horizontal motion speed ($r(1065) = -.083$, $p < .05$), and percussive APWA correlated with horizontal motion speed ($r(1065) = -.08$, $p < .05$) and angularity ($r(1065) = .063$, $p < .05$). Harmonic APWA correlated with with horizontal motion speed ($r(1065) = -.083$, $p < .05$) and size ($r(1065) = .064$, $p < .05$). Since autocorrelation is also used to find the fundamental pitch of a signal there is a possibility that it is detecting some pitch information especially when only using the harmonic components of the signal. Assuming this is true and keeping the assumption in mind that CMC research has found high pitch to be related to small size, the correlation should be negative.

3-4 Sound feature analysis multiple regression

Multiple regression analyses were run to predict the shape characteristics variables from as many sound features as possible. All sound features were introduced into the analysis and insignificant contributors were excluded from the models. Angularity was significantly predicted from from the harmonic spectral centroid averages and MFCC weighted means $F(2, 1064) = 5.390$, $p = .005$, $R^2 = .01$. Brightness was predicted from the weighted harmonic salience $F(2, 1064) = 5.431$, $p < .05$, R^2 =.005. Size was predicted from the harmonic autocorrelation and onset count $F(2, 1064) = 4.611$, $p < .05$, $R^2 = .009$. Horizontal motion speed was predicted from the harmonic autocorrelation $F(2, 1065) = 7.360$,

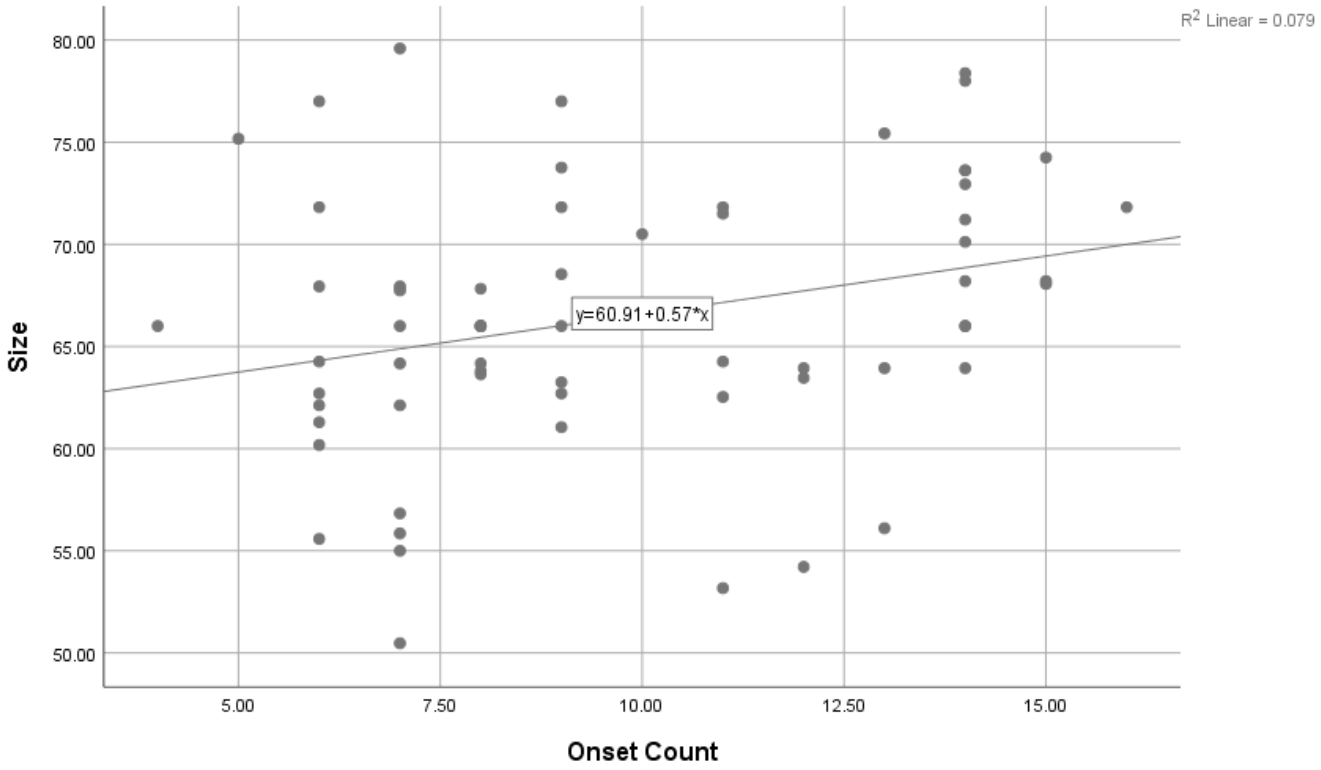

Fig. 15. Plot of shape size means and onset count values

$p < .05$, $R^2 = .007$ and horizontal motion size was predicted from the spectral centroid average, harmonic autocorrelation and the highpass magnitude average $F(3, 1063) = 5.463$; $p < .005$, $R^2 = .015$. The highpass average together with the harmonic salience also predicted the vertical motion size $F(2, 1064) = 4.419$, $p < .05$, $R^2 = .008$. Vertical motion speed, clockwise/counterclockwise motion and fade could not be significantly predicted from the sound features. Table 2 shows all the relevant multiple regression results. The R^2 value obtained from a multiple regression is a metric of how much variance in the dependent variable (angularity for example) is explained by the independent variables (harmonic spectral centroid averages and MFCC weighted means). In percentages, a R^2 value of .01 means that 1% of the total variance of the variable angularity can be accounted for by the independent variables. Possible reasons for these observations are offered in the discussion segment.

Plotting the means of the shape values and the sound features that correlate with them for every sound clip shows their relation. Figure 14 the angularity and spectral centroid averages plotted on the y-axis and x-axis respectively. Figure 15 shows the correlation between shape size and the onset count. This relation is less strong than the one in Figure 14 as can be seen in the R^2 values (.099 versus .079).

4. Discussion

This research attempted to integrate findings from CMC research into a single design. Due to oversights in this design and the challenge of extracting useful sound features from music, the research has to a great degree focused on finding relevant sound features for the chosen design of the shapes. The results show that using the features from the libROSA python library to predict all but three variables (vertical motion speed, clockwise/counterclockwise motion and fade) was successful to the degree that the experiment design allows. These results set in place a framework for predicting shape characteristics associated with sound features. Based on these findings the hypothesis drawn from CMC research that more angular, brighter shapes are associated with higher pitch is supported. The hypothesis that smaller sizes are associated with higher pitches cannot be accepted, due to the fact that it could not be predicted from the averaged pitch features. The perceptual relation between angularity and brightness and overall pitch have also been mentioned by the participants. On multiple occasions they remark that they consistently chose brighter, more spiky shapes with higher sounds and the results from the music features seem to reflect these statements. There were some unexpected results such as the high passed signal magnitudes correlating significantly with movement but not with other shape variables. Also, features that would presumably represent temporal phenomena like onset amount and autocorrelation periods did not effectively predict movement characteristics. The low but significant correlations and R^2 values are believed to result from the distribution of variable values across the shapes when they are presented to the participants. For example the likelihood of a shape that is completely congruent to

high pitch in all its characteristics (high angularity, bright and small) is very small and participants will likely make a compromises that lead to their choices containing incongruent values. As for the generative application of this project, a framework for generating new shapes has been defined. The analyses show that six out of the nine variables can be predicted from the audio features. The directions of the associations and the range for the extracted sound features can be used to define a model for predicting the shape corresponding to a sound clip (this will be demonstrated during the presentation of this thesis).

5. Limitations and further research

Multiple participants indicated that once they were familiar with the procedure they had distinct preferences for brighter and more angular shapes when the sounds were higher, which might explain the consistency of the correlation between these variables and the measures of pitch. This does also suggest that these measures do represent the perception of pitch well. Participants remarks also indicated that there were trials on which the choice of shape was directly and unequivocally certain. Unfortunately, no clear way of identifying these clips was available for analysis. For further research following this framework, a timer could be used to keep track of the amount of time that participants take to choose a shape, the shorter it is, the more certain the participants choices is. Alternatively, a confidence of choice measurement can be included. Recording the values of the shapes that were not chosen would also offer more useful data on what sound features participants do not associate shapes with. The angularity also seems to have consistently influenced the overall size of the shape. This is due to an oversight during the design phase of the project and might have contributed to a lesser ability to predict the shape from the sound features. Seeing that high pitch is congruent with small shapes and angular shapes, this oversight makes it difficult to discern the difference in predictiveness of the sound features. Motion seems to have also been an important factor to participants and one that might have not been represented in a manner that is easily quantified or accurate. Larger motion sizes and speeds affect the characteristics of the generated shape to a large degree. Despite this, angularity and brightness remain relatively well predicted by the pitch features. It remains unclear what role motion plays in these interactions, however. In future research using this framework, snapshots of the shapes state could be recorded and matched to time series data from the audio analyses. A note onset in time could be matched with a shapes position at a certain moment perhaps. In general using an approach that changes the total shape due to the horizontal and vertical motion might offer a clearer view into the predictive qualities of the sound features for each shape characteristic. This research set out to recreate findings from CMC research within a holistic design. To a certain degree, this has been achieved, with results pointing in the same direction for two of the three target variables (angularity, brightness, but not size). In the process, the foundation of how to represent sound in a meaningful way for these purposes has been created.

6. Conclusion

Music visualization typically uses a mapping between visual characteristics and musical features which are defined by the artist rather than associations found through research on the correspondences between the modes of sight and sound. This research aimed to integrate multiple of these findings into a singular design and test whether the associations between shape and sound still hold. It also aimed to explore what methods of analysing sound can be used to map these findings on to sound features. The results show that some expectations were confirmed, but not all and that there seem to be music features that relate to the visual perception of sound. Future research should aim to avoid the methodological mistakes this one has made by limiting the unpredictable range of behaviors the visualization can exhibit and use a more controlled sound for analysis. This research shows that there is a workable alternative to musical representations that are dictated by personal preference only.

References

Degani, A., Leonardi, R., Migliorati, P., & Peeters, G. (2014, September). A Pitch Salience Function Derived from Harmonic Frequency Deviations for Polyphonic Music Analysis. In DAFx (pp. 195-201).

Dubnov, S. (2004). Generalization of spectral flatness measure for non-gaussian linear processes. IEEE Signal Processing Letters, 11(8), 698-701.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. Journal of Vision, 10(1), 6:112.

Fitzgerald, D. (2010) Harmonic/Percussive Separation using Median Filtering. 13th International Conference on Digital Audio Effects (DAFX10), Graz, Austria, 2010.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. Perception & Psychophysics, 68, 11911203.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. The Journal of the Acoustical Society of America, 63(5), 1493-1500.

Marks, L. E. (1987). On cross-modal similarity: Auditoryvisual interactions in speeded discrimination. Journal of Experimental Psychology: Human Perception and Performance, 13, 384394.

Patton, K. (2007). Morphological notation for interactive electroacoustic music. Organised Sound, 12(2), 123-128.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. Journal of consciousness studies, 8(12), 3-34.

Sapir, E. (1929). A study in phonetic symbolism. Journal of Experimental Psychology, 12, 225239.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics, 73(4), 971-995.