Timbre and Visual Forms: a crossmodal study relating acoustic features and the Bouba-Kiki Effect

Ivan Eiji Simurra[1,2†], Patrícia Vanzella[2] and João Ricardo Sato[2]
[1] Music Department, Federal University of Acre, Rio Branco, Acre, Brazil
[2] Center of Mathematics, Computing and Cognition, Federal University of ABC, São Bernardo, São Paulo, Brazil
[†] Corresponding author: ieysimurra@gmail.com

Introduction

Music has a multidimensional nature with a myriad of features set over time that vary in a multitude of acoustic profiles. It has been shown, for instance, that music listening may be a multimodal experience where musical sounds can evoke abstract visual forms. Particularly, we are interested in a well-described effect known as the Bouba-Kiki effect (Köhler, 1929). This phenomenon relates to a non-arbitrary tendency to associate abstract words whose utterance demand rounding sounds (as in Bouba) with rounded shapes, while sharp words (as in Kiki*)* are usually associated with angular shapes. The studies based on the bouba-kik effect provide the first vital clues to understand the origins of proto-language, as it suggests that there may be natural restrictions on the way sounds are mapped on objects. Previous research suggests that this cross-modal phenomenon may also be found between musical timbre and shapes (Adeli et al, 2014). The authors studied the cross-modal correspondences between musical timbre and visual forms. Basically, using visual stimuli based on the literature about the Bouba-Kiki effect, each sound stimulus with timbre variation was related with some peculiarities of shapes, that is, rounded or angular. For the music orchestration and timbre studies, contemporary music, particularly from the Second Half of the 20th Century and the 21st Century music compositions, have made extensive use of technical procedures to draw attention to novel sound characterization features, such as texture and the presence of noisy sounds as relevant sound events. Such prospect is delved in the context of non-standard instrumental techniques concurrent to the usage of alternative musical orchestration settings. The present study focuses on the cross-modal correspondences between the acoustic correlates of contemporary orchestral music and the visual forms from the Bouba-Kiki Effect. We carried out an online experiment to collect ratings from subjects listening to contemporary music excerpts. Then, we cluster the classification to summarise results and then we analyze them by the acoustic features from auditory stimuli.

Method

The sound stimuli database was designed to address contemporary music techniques and practices aimed at the creation of new sounds and textures in orchestral writing (Griffths, 1978). A total of eleven sound stimuli were selected each with a duration of 5.0 second. Audio mixings were created using Audacity and the instrumental audio samples used to generate the orchestral sound textures belong to three sound databases (Ballet et al, 1999). The audio fragments were chosen from excerpts of chamber music and orchestral works by composers such as Ravel, Debussy, Stravinsky, Messiaen, Schoenberg, Ligeti, Grisey, Scelsi, Lachenmann, Xenakis and Sciarrino. Accordingly, such compositions and instrumental techniques may be explored to create unexpected sound effect modifying the global timbre perception. For the visual shapes we selected forms with different structures between rounded and jagged features. For that, we settled our assortment based on the study performed by Nielsen and Rendall (2011). On the first step of the experiment, we gathered data from an online survey in order to select the most appropriate visual shapes by listening each of the auditory stimulus. Fifty-one volunteers (30 women, age average = 34.79, sd ± 9.80) rated each auditory stimulus with one of the ten abstract shapes alternatives. To avoid fatigue effects and other biases responses both the sound events and the visual shapes were randomized presented for every subject. Moreover, participants were allowed to listen to each audio excerpt many times while they filled their ratings. The experiment lasted about nine minutes on average, according to log data retrieve from the online experiment. Figure 1 depicts the online interface.

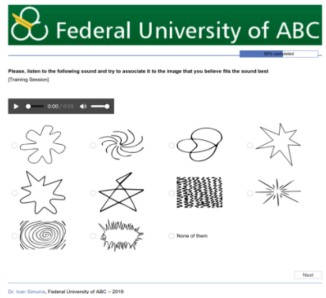

Figure 1: online interface for each auditory stimulus.

We then applied K-means to cluster the visual forms according to the subject ratings. A total of 4 groups of clusters was achieved. Chi-squared tests were conducted to determine whether shapes had been selected randomly, with results suggesting that shape selection was very unlikely to have arisen by chance: cluster 1, $p < 2.2e\text{-}16$; 2, $p = 9.365e\text{-}09$; 3, $p = 1.202e\text{-}04$ and 4, $p = 4.086e\text{-}06$. Figure 2 depicts the 4 most consistent visual forms selected based on the K-means results.

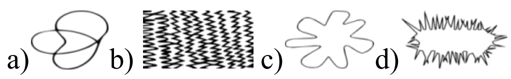

Figure 2: The four visual forms selected from K-means cluster techniques

Additionally, we also grouped each of the audio stimuli presented by the K-means indexes. Figure 3 displays all the audio stimuli according to the K-means plot visualization.

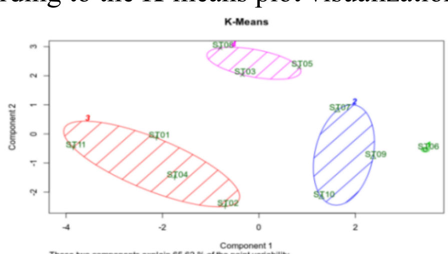

Figure 3: K-means plot featuring all the 11 audio stimuli

Next, we examined the acoustic features of the audio stimuli presented to participants in order to determine any relationship with the visual forms selected. For that, we used Sonic Visualiser (Cannam et. al, 2010) application with specific function libraries to the following acoustic features processing: Spectral Centroid, Spectral Standard Deviation, Spectral Kurtosis, Spectral Flatness, Spectral Flux, Spectral Irregularity, Odd to Even Ratio, Zero Crossing Rate, Energy RMS and Spectral Rolloff (Bullock, 2007). Therefore, for each of the 11 audio stimuli, we retrieved a vector of 10 acoustic features. We then analyzed the acoustic data to examine the presence of multicollinearity among the different features. We assume that on the used acoustic features analysis vectors there might be a some similarity among them. This can result in artifacts data and consequently in more noisy results. To verify if it is occurs we analyzed the multicollinearity. For that, we performed the Kaiser-Meyer-Olkin (KMO) model to predict whether data are likely to factor well, based on correlation and partial correlation (Zacharakis et al, 2014). First, all variables were normalized from the whole 11 audio stimuli between range [0,1]. Then we performed a KMO test to verify collinearity. On the acoustic features vector, we applied Principal Components Analysis (PCA) to reduce dimensionality from variables. Finally, we calculated Factor Analysis (FA) to determine the most expressive acoustic features by each component. Next Section evinces results from the method described.

Results

Each of the acoustic features retrieved from each sound stimulus resulted in a overal KMO Measure of Sampling Adequacy (MSA) index greater than 0.69 (0.76, 0.85, 0.78, 0.69, 0.75, 0.79, 0.78, 0.76, 0.84, 0.73, 0.85). According to Zacharakis, KMO overall should be .60 or higher to proceed with factor analysis. The average of the overall KMO for the 11 sound events was 0.76. Following Henry Kaiser evaluation (1974) this result in middling and its necessary to discuss it with caution. After, we calculated PCA to determine the number of components for the FA (rotation: 'varimax' and score: 'Bartlett'). We defined the number of 04 principal components which explain a total average of 88% of the cumulative proportion of the data variance. The prominent descriptors over the four factors for each groupings by visual shapes are shown in Tables 1.

Table 1: Factor loadings by Figures 1a, 1b, 1c and 1d.

	Figure 1a				**Figure 1b**			
Acoutic Feature	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**
SpecCentroid		*0.850*	*0.504*	*0.128*	*0.908*	*-0.114*	*-0.204*	*-0.285*
SpecFlatness	*-0.503*		*0.858*		*-0.310*	*0.183*	*0.913*	*0.179*
SpecFlux		*0.203*		*0.640*	*0.519*	*0.340*	*-0.170*	
SpecIrregularity	*0.970*			*-0.113*	*0.450*	*0.808*	*-0.107*	*0.194*
Odd to Even	*0.377*	*-0.178*	*0.231*	*0.181*		*-0.117*	*0.145*	*-0.127*
SpecStandardDev	*-0.243*	*0.615*	*0.406*	*0.524*	*0.963*		*-0.147*	
Zero Crossing	*0.113*	*0.263*	*0.595*		*0.379*	*-0.102*	*0.118*	*-0.545*
RMS	*0.985*					*0.677*		*0.650*
SpecKurtosis	*0.907*	*-0.130*	*0.338*			*0.921*	*0.153*	*0.124*
SpecRoll Off		*0.879*			*0.916*	*0.286*	*-0.105*	

	Figure 1c				**Figure 1d**			
Acoutic Feature	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**
SpecCentroid	*0.893*	*-0.266*	*-0.280*	*-0.176*	*-0.286*	*0.929*	*0.123*	*-0.189*
SpecFlatness	*-0.233*		*0.581*	*0.578*	*-0.913*	*0.337*		
SpecFlux		*0.296*	*0.587*	*0.188*	*0.140*		*0.480*	*0.347*
SpecIrregularity	*-0.115*	*0.818*	*0.402*		*0.863*	*-0.231*		
Odd to Even		*0.111*	*0.427*				*0.471*	
SpecStandardDev	*0.979*	*-0.168*			*-0.302*	*0.950*		
Zero Crossing	*0.580*	*-0.343*	*-0.388*	*-0.546*	*-0.363*	*0.264*	*0.699*	
RMS	*-0.286*	*0.921*	*0.168*	*0.191*	*0.827*	*-0.268*	*-0.463*	
SpecKurtosis	*-0.628*	*0.433*	*0.260*	*0.218*	*0.758*	*-0.469*		*0.151*
SpecRoll Off	*0.948*		*0.139*		*-0.341*	*0.891*		

Discussion

The results indicated that, irrespective of the contrasting visual shape attributes, the prevalence of the spectral magnitude, mainly associated with noise content and the specific magnitude on spectrum region, is a substantial acoustic feature for all four groupings. However, some peculiarities inherent to the sound qualities associated with each figure were observed. For jagged and sharp forms, Spectral Kurtosis and Spectral Flatness could be highlighted as the acoustic features that were most associated with angular shapes. Interestingly, previous findings suggest that these acoustic features are generally associated with noise sounds (Peeters, 2003; Krimphoff et al, 1994; Simurra and Manzolli, 2016). On the other hand, the most prominent acoustic features associated with rounded and smooth shapes were Energy RMS and the Spectral Roll-Off, which are often related to the brightness of the sound (Peeters, 2003; Krimphoff et al, 1994; Simurra and Manzolli, 2016). These results thus suggest that certain symbolic visual features seem to share acoustic resources at some stage, such as those centered on noisy sound. Moreover, by using only spectral features content with only a small portion of spectro-temporal features results centered only on the spectral magnitude output. Some other audio descriptors will be necessary to test spectro-temporal and temporal features effect. It is interesting to pinpoint for contemporary music repertoire that presents, among multiple variables, non-harmonic spectral content. Further studies would be necessary to compare or correlate groupings sharing similar features.

Acknowledgments

The authors gratefully acknowledge financial support from (CAPES-88887.341537/2019-00). We thank Thenille Braun Janzen for her suggestions on earlier versions of the manuscript.

References

Adeli, M., Rouat, J., & Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Frontiers in human neuroscience, 8*, 352.

Ballet, G., Borghesi, R., Hoffmann, P., & Lévy, F. (1999). Studio online 3.0: An internet" killer application" for remote access to IRCAM sounds and processing tools. *Journées d'informatique musicale (JIM).* (pp. 123–131). Issy-les-Moulineaux, France.

Bullock, J. (2007). Libxtract: A Lightweight Library for audio Feature Extraction. Conservatoire. In *Proceedings of the International Computer Music Conference (ICMC)*, *43,* 25-28.

Cannam, C., Landone, C., & Sandler, M. (2010, October). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. ACM

Griffiths, P. (1978). *A concise history of avant-garde music: from Debussy to Boulez*. New York: Oxford University Press.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36.

Köhler, W. (1929). *Gestalt psychology. Psychological research*, Springer, v. 31, n. 1, p.XVIII–XXX

Krimphoff, J., Mc Adams, S., and Winsberg, S. (1994). Caractérisation du timbre dessons complexes. ii. analyses acoustiques et quantification psychophysique. *Le Journal de Physique IV*, vol. 4.

Nielsen, A., & Rendall, D. (2011). The sound of round: evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology,* 115.

Peeters, G. (2004). A large set of audio features for sound description in the cuidado project. *IRCAM*.

Simurra, I. E., Manzolli, J. (2016). Sound Shizuku Composition: a Computer-Aided Composition Systems for Extended Music Techniques. *Brazilian Journal of Music and Mathematics*, 86-101.

Zacharakis, A. I., Pastiadis, K., Papadelis, G., & Reiss, J. D. (2011). An Investigation of Musical Timbre:Uncovering Salient Semantic Descriptors and Perceptual Dimensions. In *ISMIR,* 807-812.

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception*, *31*(4), 339-358.