Which Timbral Features Granger-Cause Colour Associations to Music?

PerMagnus Lindborg
School of Creative Media, City University of Hong Kong, HKSAR
pm.lindborg@cityu.edu.hk

Introduction

Sensory information processing is inherently multimodal. An organism normally perceives the environment using all its senses simultaneously. Crossmodal correspondence might take place at any stage of neural processing (Spence 2011; Deroy & Spence 2016), and studies have provided evidence that many non-arbitrary correspondences exist between auditory and visual stimulus features (Martino & Marks 2001; Bresin 2005; Palmer et al. 2013; Whiteford et al. 2018). However, few studies take electroacoustic music compositions as substrate for stimuli, despite the great variations of timbre within and across such works. This paper outlines an ongoing study of audiovisual correspondences through time series analysis. We investigate Granger-causality and other measures of association from time series analysis of multivariable acoustic features (describing nine music excerpts via timbral and dynamic features) and multivariate visual features (size and colour), collected in a perceptual experiment (N = 21) with a continuous response interface.

In previous work we charted colour-to-sound associations in film music (Lindborg & Friberg 2015) and in electroacoustic music (Lindborg, 2019a) with audio excerpts of around 15 seconds. Since the stimuli were fairly short they could be treated as independent data points in a random distributed variable. The present study extends these studies to longer music excerpts through time series analysis techniques. Listening to a recorded piece does not alter the audio file and in this sense the information flow is one-directional. Hence for our purposes the acoustic features may be treated as independent variables and colour responses as dependent variables in a time series regression analysis. A paper detailing experiments on continuous responses to music was published by Emery Schubert (1999) who then extended the analysis to time series (2001). Recent developments are in no small measure due to Roger Dean and collaborators (e.g. Dean & Bailes, 2010; 2011; Pearce ,2011; Bailes & Dean, 2012; Dean & Dunsmuir, 2016).

Materials

Audio excerpts of approximately three minutes duration were selected from nine electroacoustic pieces: well-known works by Chowning, Harvey, Risset, and Wishart, as well as recent pieces by Winderen, Martin, and the author. After normalisation by loudness (Nygren, 2009), they were presented in randomised order in an experiment (N = 21; eight females, median age 30, all right-handed, no reported colour vision deficiency or hearing impairment) following the same procedure as in (Lindborg, 2019a). While listening, participants manipulated two interfaces with the hands to control the size and colour of a visual object presented on a screen. Their task was to continuously match this object to the music.

Responses were sampled at 10 Hz, and colour was represented in *CIELab* which closely matches human perception (Hoffman, 2003; Shaw & Fairchild, 2002). It has three orthogonal dimensions that correspond to lightness (L), green-to-red (a), and blue-to-yellow (b). See example in Figure 1, left panel. Specifying colours within a perceptual scheme has advantages over parametric schemes (such as RGB or HSL) in terms of replicability and relevance to visual perception. Colour spaces and the design of the experimental response interface are discussed in (Lindborg & Friberg, 2015).

A large number of acoustic features were extracted computationally using the *MIR Toolbox* (Lartillot, 2013). Note that most are highly inter-correlated, due to the way the algorithms are structured. A selection of around 20 was made based on previously reported results, and to these we joined psychoacoustic descriptors extracted with *PsySound* (Cabrera et al., 2008). Time series were cleaned by imputing missing values (0.03% of responses in total) and handling outliers (altering 0.3% of the most extreme values two-

tailed, corresponding to trimming at ±3 SD in a normal distribution), and all series were down-sampled to a common rate of 4 Hz.

Methods

In empirical time series the elements almost always display some form of serial dependency. In our experimental setup, it is clear that if the music changes the visual response object is likely to be changed as well, but it will do so gradually, since the new position of the interfaces depend on their previous positions. Each new data point is to some degree correlated with the preceding ones: the time series is autocorrelated. In order to use parametric statistics to evaluate the degree of association between two time series, the values must be independent and identically distributed within each.

We conducted analysis in *R* (R Core Team, 2020) following the approach outlined in (Dean & Dunsmuir, 2016), with each music excerpt considered as a separate case study. For details on the statistical methods mentioned below see e.g. (Hyndman & Athanasopoulos, 2018) especially chapter 8, and (Box et al., 2015), especially chapters 4–5. See also (Jebb et al., 2015) for applications in psychology, and (Pfaff, 2008) for econometrics.

Time series were reshaped to achieve 'weak stationarity' by differencing. This removes trends in the data. In nearly all cases, d = 1 was an adequate degree, as judged by the KPSS and Augmented Dickey-Fuller tests. We then performed initial modelling tests including Granger causality (Granger, 1969) as an exploratory tool to assess whether the relationship between two stationarized series contains a causal element, at some lag. See Figure 1, right panel, for an example. However, it does not inform us about the strength of the predictive causation nor yield a transfer function that allows us to model the relationship.

In predictive regression modelling we need to be able to evaluate the cross-correlations between series, at a range of lags. Before significance levels of predictors can be correctly assessed the autocorrelation needs to be removed from at least one of the series being compared. This 'prewhitening' process involves modelling the autocorrelation structure so that the residuals display desired statistical properties, including serial independence, normal distribution, and heteroskedasticity. For each series, we estimated the autoregressive components with an ARIMA model. After obtaining reasonable maximum parameters (for p and q, since d was previously determined) from the autocorrelation and partial autocorrelation functions, we searched from one degree higher (p_{max}+1, d+1, q_{ma}x+1) down to (0, 0, 0) and then selected an optimal solution, as indicated by BIC, among those where the residuals passed the portmanteau Ljung-Box test on a range of lags. However, a fully automatic process might lead to overfit i.e. models that do not generalise well. Ultimately, our primary interest is mechanistic, seeking to identify psychological mechanisms by which crossmodal association processes might be explained. Robust predictive modelling is an important step towards this goal.

Therefore we are currently investigating the *auto.arima()* function (Hyndman et al., 2018) which implements a more powerful search method and an interface that is flexible when multiple exogenous predictors are included, i.e. ARIMAX. Since several acoustic features can potentially be influencing the colour response, a parsimonious set of predictors can be found by a systematic process of stepwise reduction, where predictor coefficient significance, error variance, and BIC are used as guides as to which predictors to include or exclude. For each case under study, the resulting model informs us of the predictive influence onto a dependent response variable from three sources: its own autoregressive function, a transfer function of the optimal set of acoustic predictor variables at different lags, and a white-noise error term. The transfer function is our focus of interest. The explanatory strength of the model is estimated from the cross-correlation series over a range of lags.

In our data the response is a multivariate time series {Size, L, a, b}. We are currently investigating univariate series separately, i.e. Size and a Change variable derived from the four, both for individual participants and for a group average (see example in Fig. 1). As the *CIELab* variables display multicollinearity, a full analysis requires a multivariate approach.

Discussion

The modelling approach outlined above assumes that the character of crossmodal associations at play are stable over time. It is an acceptable simplification given the experiment at hand, but the analysis of a natural situation calls for a dynamic approach. We have assumed that in the course of the visual association task, the listener "locks in" on something that s/he hears, and chooses a strategy, most likely intuitively, to match the colour response. As shown in (Lindborg & Friberg, 2015), emotion can be a strongly mediating factor. When presented with another stimulus, it might be that another acoustic feature takes on greater salience and a therefore a different matching strategy emerges. During the time the response is "locked in", a higher degree of influence of one or more acoustic features onto one or more of the response parameters will be observed.

The continuous response method supports situations where expert subjects are closely tracking their perception of spectro-morphological features of music. Using colour is a way to side-step semantic cognitive processing and can potentially reflect lower-level crossmodal processing mechanisms. Such methods can provide advantageous experimental tasks with non-expert subjects, or for those unable to use words (such as small children or stroke patients), or in experimental situations where the cognitive load of having to translate perceptions into semantic labels might distract from the task or from the act of listening itself (Lindborg, 2019b; cf. Saitis et al., 2019).

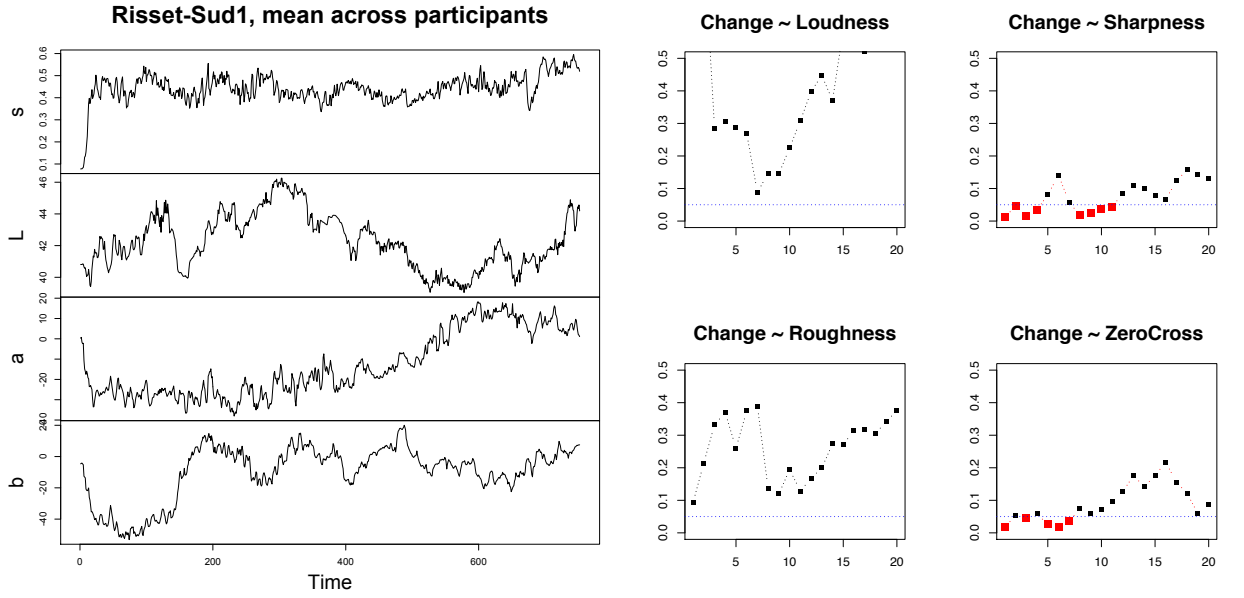

Figure 1: Plots for J.-C. Risset's Sud (three-minute excerpt from the beginning of the first part). Left: Size, L, a, b time series averaged across participants. Right: Change response Granger-caused by respectively Loudness, Sharpness, Roughness, and ZeroCross for a range of lags. Squares (red colour) below the dotted line indicate lags with a predictive-causal relationship significant at α = 0.05.

References

Bailes, F., & Dean, R. T. (2012). Comparative time series analysis of perceptual responses to electroacoustic music. *Music Perception: An Interdisciplinary Journal, 29*(4), 359-375.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*: John Wiley & Sons.

Bresin, R. (2005). What is the color of that music performance? *Proceedings of the International Computer Music Conference*

Cabrera, D., Ferguson, S., Rizwi, F., & Schubert, E. (2008). PsySound3: a program for the analysis of sound recordings. *Journal of the Acoustical Society of America, 123*(5), 3247.

Dean, R. T., & Bailes, F. (2010, Oct.). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review, 5:4*, 152-175.

Dean, R. T., & Bailes, F. (2011, Apr.). Modelling perception of structure and affect in music: Spectral centroid and Wishart's Red Bird. *Empirical Musicology Review, 6:2*, 131-137.

Dean, R. T., & Dunsmuir, W. T. (2016). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior research methods, 48*(2), 783-802.

Deroy, O., & Spence, C. (2016). Crossmodal correspondences: Four challenges. *Multisensory research, 29*(1-3), 29-48.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.

Hoffmann, G. (2003). *Cielab color space*. Retrieved from http://docs-hoffmann.de/cielab03022003.pdf

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., . . . Yasmeen, F. (2018). forecast: Forecasting functions for time series and linear models. R package version 8.4. *URL: https://CRAN. R-project. org/package= forecast*.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*: OTexts.

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: examining and forecasting change. *Frontiers in psychology, 6*, 727.

Lartillot, O. (2013). MIRtoolbox 1.5, User's Manual. *Finnish Centre of Excellence in Interdisciplinary Music Research*.

Lindborg, P. (2019a.). What is the Color of that Electroacoustic Music? *Proceedings of the International Computer Music Conference j.w. New York City Electroacoustic Music Festival*, New York, NY, USA.

Lindborg, P. (2019b). How do we listen? 에밀레 *Emille Journal of the Korean Electro-Acoustic Society, 16*, 43-49.

Lindborg, P., & Friberg, A. K. (2015). Colour association with music is mediated by emotion: Evidence from an experiment using a CIE Lab interface and interviews. *PloS One, 10*(12).

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science, 10*(2), 61-65.

Nygren, P. (2009). *Matlab code for the ITU-R BS. 1770–1 implementation. Appendix E to Master Thesis: Achieving Equal Loudness between Audio Files.* (MSc). KTH Royal Institute of Technology,

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences, 110*(22), 8836-8841.

Pearce, M. T. (2011, Apr.). Time-series analysis of music: Perceptual and information dynamics. *Empirical Musicology Review, 6:2*, 125-130.

Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R*: Springer Science & Business Media.

R Core Team. (2020). R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Saitis, C., & Weinzierl, S. (2019). The semantics of timbre. In *Timbre: Acoustics, perception, and cognition* (pp. 119-149): Springer.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Series in affective science. Music and emotion: Theory and research* (pp. 393–414): Oxford University Press.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology, 51*(3), 154-165.

Shaw, M., & Fairchild, M. (2002). Evaluating the 1931 CIE color-matching functions. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, 27*(5), 316-329.

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

Whiteford, K. L., Schloss, K. B., Helwig, N. E., & Palmer, S. E. (2018). Color, music, and emotion: Bach to the blues. *i-Perception, 9*(6), 2041669518808535.