Fachgebiet Audiokommunikation
Prof. Dr. Stefan Weinzierl

Master Thesis

Modelling Timbre Similarity of Long Solo Instrument Clips

The objective of this project is to quantitatively model human perceptual similarity responses to musical sounds embedded in richer temporal, melodic, harmonic, and dynamic contexts using low-level properties of the audio signal. Constructing a cognitively adequate model of timbre similarity is useful for understanding what timbral cues listeners extract and process from long-duration and context-rich music audio clips.

Previous work on modelling timbre similarity between orchestral instrument sounds has concentrated on the processing and analysis of temporally isolated notes [1]. However, a representation of musical timbre is only useful to the extent that it accounts for changes in musical context (i.e., horizontal and vertical combinations of pitches, dynamics, durations, and articulations) and for particularities between different types of musical sound (e.g., sustained versus impulsive, string-bowed versus wind-blown). As an example, it has been shown that the importance of temporal cues, such as attack time and slope, is greatly reduced in melodically or harmonically combined versus isolated notes [2]. Successful efforts to model the perceived similarity of short music clips with low-level acoustic timbre descriptors have demonstrated the potential for applying music information retrieval techniques to model complex perceptual processes [3], [4].

The modelling approach will be based on a latent-variable linear regression technique, namely partial least-squares regression (PLSR) [4]. First, pairwise similarity ratings of solo instrument excerpts from recorded music will be obtained from listeners. Perceptual similarity distances will be used as target variables in the PLSR model. Subsequently, audio descriptors will be extracted from the sound clips [5], [6] and their pairwise similarities will be used as predictor variables. Given the nature of the sound stimuli to be used, it is of special interest to determine timbre descriptors or representations of descriptors that remain relatively invariant over note, harmonic, and dynamic transitions in an instrument.

Literature

[1] K. Siedenburg, K. Jones-Mollerup, and S. McAdams, “Acoustic and categorical dissimilarity of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds,” *Frontiers in Psychology*, vol. 6, p. 1977, 2016.

[2] J. M. Hajda, “The effect of dynamic acoustical features on musical timbre,” in *Analysis, Synthesis and Perception of Musical Sounds*, J. W. Beauchamp, Ed., New York: Springer, 2007, pp. 250–271.

[3] R. Ferrer and T. Eerola, “Semantic structures of timbre emerging from social and acoustic descriptions of music,” *EURASIP J. Audio Speech Music Process.*, 2011.

[4] K. Siedenburg and D. Müllensiefen, “Modeling timbre similarity of short music clips,” *Frontiers in Psychology*, vol. 8, 2017.

[5] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, “The Timbre Toolbox: Extracting audio descriptors from musical signals,” *J. Acoust. Soc. Am.*, vol. 130, no. 5, pp. 2902–2916, 2011.

[6] O. Lartillot and P. Toiviainen, “A matlab toolbox for musical feature extraction from audio,” in *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, 2007, pp. 237–244.

Requirements

- Well grounded knowledge in audio signal processing
- Solid skills in MATLAB or Python programming
- Experience with audio recording and editing

Supervisors

Dr. Charalampos Saitis	H 3001A	charalampos.saitis@campus.tu-berlin.de
Prof. Dr. Stefan Weinzierl	E-N 322	stefan.weinzierl@tu-berlin.de

Berlin, March 22, 2018