EXTENDED PLAYING TECHNIQUES: THE NEXT FRONTIER IN MUSICAL INSTRUMENT RECOGNITION

First Author
Affiliation1
`author1@ismir.edu`

Second Author
Retain these fake authors in submission to preserve the formatting

Third Author
Affiliation3
`author3@ismir.edu`

ABSTRACT

Although the automatic recognition of a musical instrument from the recording of a single "ordinary" note is close to becoming solved problem, the ability of a computer to precisely identify playing technique within an extended taxonomy remains far below human accuracy. This article provides the first benchmark of machine listening systems for the classification of extended playing techniques in the symphonic orchestra. We identify three necessary conditions for significantly outperforming the classical mel-frequency cepstral coefficients (MFCC) baseline: (1) the inclusion of second-order scattering coefficients to account for the presence of amplitude modulations ; (2) the inclusion of large temporal scales ; and (3) the resort to supervised metric learning.

1. INTRODUCTION

The progressive diversification of the timbral palette in Western classical music at the turn of the 20th century is reflected in five concurrent trends: the addition of new instruments to the Western symphonic instrumentarium, either by technological inventions (e.g. theremin) or importation from non-Western musical cultures (e.g. marimba) [12]; the creation of novel instrumental associations, as epitomized by *Klangfarbenmelodie* [13]; the temporary alteration of resonant properties through mutes and other "preparations" [3]; a more systematic usage of extended instrumental techniques, such as artificial harmonics, *col legno batutto*, or flutter tonguing [7]; and the resort to digital audio effects [17]. The first of these trends has somewhat stalled: to this day, most Western composers rely on an acoustic instrumentarium that is only marginally different from the one that was available in the Late Romantic period. Nevertheless, the latter approaches to timbral diversification were massively adopted into post-war contemporary music. In particular, an increased concern for the concept of musical gesture [5] has liberated many unconventional instrumental techniques from their figurativistic connotations, thus making the so-called "ordinary" playing style merely one of many compositional options.

Far from being exclusive to erudite music, extended playing techniques are also commonly found in oral tradition; in some cases, they even stand out as a distinctive component of musical style. Four well-known examples are: the snap pizzicato ("slap") of the upright bass in rockabilly, the growl of the tenor saxophone in rock'n'roll, the shuffle stroke of the violin ("fiddle") in Irish folklore, and the glissando of the clarinet in Klezmer music. Consequently, the mere knowledge of organology (the instrumental *what?* of music), as opposed to chironomics (its gestural *how?*), is a rather weak source of information for browsing and recommending music within large audio databases.

Yet, past research in music information retrieval (MIR), and especially machine listening, too rarely acknowledges the benefits of integrating the influence of performer gestures into a coherent taxonomy of musical instrument sounds. Instead, gestures are either framed as a spurious form of intra-class variability between instruments, without delving into its interdependencies with pitch and intensity; or, symmetrically, as a probe for the acoustical study of a given instrument, without enough emphasis onto the broader picture of orchestral diversity.

One major cause of this gap in research is the difficulty of collecting and annotating data for contemporary instrumental techniques. Fortunately, such obstacle has recently been overcome, owing to the creation of databases of instrumental samples in a perspective of spectralist music orchestration [10]. In this article, we capitalize on the availability of data to formulate a new line of research in MIR, namely the joint retrieval of organological information ("*what* instrument is being played in this recordings?") and chironomical information ("*how* is the musician producing sound?"), while remaining invariant to other factors of variability, which are deliberately regarded as contextual: where, when, why, by whom, and for whom was the music (in this recording) played.

2. RELATED WORK

Timbre classification of a single musical instrument (clarinet): [9]. Retrieval of percussion gestures using timbre classification techniques: [15]. Polyphonic instrument recognition using spectral clustering: [11]. Knowledge representation issues in musical instrument ontology design: [6]. Guitar playing technique classification: [14].

MedleyDB: [2]. Audio Set: [4]. Visipedia: [1] Scattering transforms in musical instrument recognition: [8, 16].

3. TASKS

4. METHODS

4.1 Scattering transform

5. LARGE-MARGIN NEAREST NEIGHBORS

6. EXPERIMENTAL RESULTS

7. CONCLUSION

8. REFERENCES

[1] Serge Belongie and Pietro Perona. Visipedia circa 2015. *Pattern Recognition Letters*, 72:15 – 24, 2016.

[2] Rachel M Bittner, Justin Salamon, Mike Tierney, Matthias Mauch, Chris Cannam, and Juan Pablo Bello. Medleydb: A multitrack dataset for annotation-intensive mir research. In *Proc. ISMIR*, 2014.

[3] Tzenka Dianova. *John Cage's Prepared Piano: The Nuts and Bolts*. PhD thesis, U. Auckland, 2007.

[4] Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. Audio set: An ontology and human-labeled dataset for audio events. In *Proc. IEEE ICASSP*, 2017.

[5] R.I. Godøy and M. Leman. *Musical Gestures: Sound, Movement, and Meaning*. Taylor & Francis, 2009.

[6] Sefki Kolozali, Mathieu Barthet, György Fazekas, and Mark B Sandler. Knowledge representation issues in musical instrument ontology design. In *Proc. ISMIR*, pages 465–470, 2011.

[7] Stefan Kostka. *Materials and Techniques of Post Tonal Music*. Taylor & Francis, 2016.

[8] Vincent Lostanlen. *Convolutional operators in the time-frequency domain*. PhD thesis, 'Ecole normale supérieure, 2017.

[9] Mauricio A Loureiro, Hugo Bastos de Paula, and Hani C Yehia. Timbre classification of a single musical instrument. In *Proc. ISMIR*, 2004.

[10] Yan Maresz. On computer-assisted orchestration. *Contemp. Mus. Rev.*, 32(1):99–109, 2013.

[11] Luis Gustavo Martins, Juan José Burred, George Tzanetakis, and Mathieu Lagrange. Polyphonic instrument recognition using spectral clustering. In *Proc. ISMIR*, 2007.

[12] Curt Sachs. *The History of Musical Instruments*. Dover Books on Music. Dover Publications, 2012.

[13] Arnold Schoenberg. *Theory of Harmony*. University of California, 100th anniversary edition edition, 2010.

[14] Li Su, Li-Fan Yu, and Yi-Hsuan Yang. Sparse cepstral, phase codes for guitar playing technique classification. In *Proc. ISMIR*, 2014.

[15] Adam R Tindale, Ajay Kapur, George Tzanetakis, and Ichiro Fujinaga. Retrieval of percussion gestures using timbre classification techniques. In *Proc. ISMIR*, 2004.

[16] Steven K Tjoa and KJ Ray Liu. Musical instrument recognition using biologically inspired filtering of temporal dictionary atoms. In *Proc. ISMIR*, 2010.

[17] Udo Zölzer. *DAFX: Digital Audio Effects*. Wiley, 2011.