Proceedings Paper:
Beeston, A.V. and Summers, M.A.C. (2013) Groundwork for a resource in computational hearing for extended string techniques. In: Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR). 10th International Symposium on Computer Music Multidisciplinary Research (CMMR), 15-18 Oct 2013, Marseilles, France. Publications du L.MA. , pp. 662-669. ISBN: 978-2-909669-23-6.

Groundwork for a Resource in Computational Hearing for Extended String Techniques

Amy V. Beeston[1] and Mark A. C. Summers[2]

[1] Department of Computer Science, University of Sheffield, UK
[2] Department of Music, University of Sheffield, UK
{a.beeston,m.summers}@sheffield.ac.uk

Abstract. Extended techniques and signal processing devices are increasingly common in contemporary music composition and performance. At present, few machine listening methods deal reliably with extended techniques. Moreover, existing instrumental corpora have not traditionally included sources of variation that arise naturally in every-day performance environments. In the current study, timbral descriptors are extracted for a range of instrumental techniques, and their dispersion is quantified in order to examine the variation stemming from recording strategy choice and performer repetition.

Keywords: Extended techniques, performance variability, instrumental dataset, timbral analysis, machine listening.

1 Introduction

As 'extended techniques' have become mainstream in contemporary music composition and performance, work is underway to develop machine listening methods that allow a more fruitful integration of acoustic instrument with signal processing technologies (see e.g., [1–5]). The current paper describes ongoing work developing a database of extended string techniques recorded under naturalistic performance conditions, and examines computational hearing methods that attempt to summarise sonically interesting aspects of sound. To motivate this work, Section 1 describes existing datasets and reveals the gap between these and modern musicians' requirements. Section 2 outlines our corpus creation methods, describing the selection of extended techniques, recording conditions, data preparation, annotation and assessment processes. Illustrative examples are presented in Section 3, and the paper concludes with a discussion in Section 4.

Extended techniques have been adopted by composers since at least the seventeenth century, and their use has become firmly established in the twentieth century as composers have sought to find new sounds on acoustic instruments. In an early example, Biber used *col legno* (the wood of the bow) and insertion of paper between the strings [6]. Krzysztof Penderecki and George Crumb worked extensively with extended string techniques in the 1960s and '70s respectively, formalising techniques far-removed from the equally-tempered chromatic scale, concerned more with timbal variation than with pitch. Extended techniques include ways of preparing and playing an instrument that fall outside traditional standard practice [7], [8] and can involve timbral changes (e.g., *sul ponticello*), unconventional musical material (e.g., *seagull glissando*), and methods to deliver sound that the instrument was not originally designed to produce (e.g., *subharmonics* [9]).

Extended techniques can be problematic for a composer to notate, and hard for a performer to achieve consistently. Moreover, the variability in reproduction of such techniques can be a serious concern for performers who use live signal processing in their work. Our auditory systems afford a degree of context-sensitivity to human hearing (e.g., constancy for spectral envelope 'colour' [10], or for the temporal envelope in reverberation [11]), but these effects have not yet been adequately replicated for machine listeners. Thus while people tend to hear the (context-dependent) 'interesting' variation in a signal, machines often end up overly reliant on cues that vary unintentionally.

Datasets containing sound material to address the variability inherent in instrumental performance (intentional or otherwise) are not common at present. Instrument corpora first documented standard performance techniques, often based around a collection of semitone-spaced pitches, in a normal working range, at a range of loudness levels [12–14]. Usefully, [12] is being updated to include second recordings of instruments, and to augment the corpus with extended techniques. A number of commercial ventures have also recognised the increasing desire for extended performance techniques [15–17], but such datasets suffer from proprietary interfaces and a lack of information regarding the audio signal processing undertaken [18]. Moreover, while such sound libraries may allow a composer to work with sound in a convenient manner, they are of no immediate use to an instrumentalist. Sample banks intended for use by performers are typically work-specific and designed to return material in real-time, for example using parameters derived from an input audio signal [1], [5].

The current study is a step towards a robust machine listener for live instrumental performance. Such a device would identify aspects of variation in sound signals that were intended, relating directly to the performance underway. Additionally, it would 'compensate' for undesired aspects of variation, resulting from uncontrollable factors such as the microphone or room effect. Towards this aim, we examine factors governing variability from a performer's point of view, and are specifically concerned with (i) performance environment and (ii) repetition. Knowledge of the expected variation in (i) would allow the performer to travel with a degree of confidence to perform in a new venue. Knowledge of the expected variation in (ii) would ensure that the performer can play the same technique several times and predict how their sound will be registered by the machine (which may allow the performer to auralise subsequent signal effect chains). Though it is outside the scope of the present paper, we note for future work that variation due to the performer and instrument themselves should be examined in order that a composer can give a performance system to different performers and achieve predictable results.

2 Methods

In the pilot study described below, a prototype corpus was used to examine the variation naturally arising in normal and extended performance techniques due to (i) the recording conditions and (ii) iteration of the technique by the performer. This section describes four main operations undertaken to gather data appropriate to the task: selection of performance techniques; selection of microphones and their placement; sample extraction and storage; automatic annotation with timbral descriptors.

2.1 Selection of Techniques

The current study draws its sound material from an ongoing project documenting the sound world of the viola da gamba. An instrument-specific list of techniques (normal and extended) has been compiled, informed by the performing background of one of the present authors (MS) with cross-reference to other surveys of extended techniques on string instruments [7], [8], [17]. A list of 90 individual techniques serves as the basis for the corpus.

A small number of these techniques have been picked for illustrative analyses in Section 3. Firstly, we fix the pitch, loudness and duration (as in typical timbre studies), and examine bowing this pitch normally on six different strings. Secondly, we use a single string to examine the effect of different bowing techniques.

2.2 Selection of Microphones and their Placement

Recordings were made in an acoustically isolated room in the University of Sheffield Sound Studios (volume 34.7 m^3). Two walls were covered with heavy felt curtains, and there was an upright piano on another wall. The player sat in one corner pointing diagonally towards a 'far' room microphone at a distance of 3.6 meters. Three further 'close' microphones were placed on or near the instrument as described in Table 1.

The signal arriving at each microphone was recorded via an RME Fireface 800 audio interface connected to a MacBook in an adjoining control studio, running Audacity software [19]. Two DPA microphones were directly attached to the instrument itself, and represent the highest signal-to-noise ratio practically

Table 1. Description of microphones selected, their directional characteristics and placement in regard to the instrument and room.

Microphone	Direction	Proximity	Placement
DPA 4060	omni	close	below bridge, under highest (1st) string
DPA 4060	omni	close	below bridge, under middle (4th) string
Neumann KM184	cardioid	close	0.1 m in front of instrument's bridge
Neumann KM184	cardioid	far	3.6 m distant to front, raised 1.8 m

achievable. The close Neumann microphone recording represents the best data capture available for a player who is unwilling to attach 'gadgets' directly to their instrument. On the other hand, the far Neumann microphone registers the signal after transformation by a small room with a moderate level of reflections, and is more representative of the sound typically transferred to the audience.

2.3 Sample Extraction and Storage

Groups of samples were recorded simultaneously with four microphones into independent channels of a long sound file. A click track of 120 beats per minute was provided over headphones to the performer (MS) to facilitate the session timing. Four seconds elapsed between the start of each sound event (2-seconds of bowing, 2-seconds of rest), and multiple iterations of every technique were recorded (six repetitions at minimum). For some techniques, the natural resonance of the instrument persisted beyond the 2-second rest period assigned, but we did not extend the recording period in such instances since it would be rare in performance to wait for the resonance to decay fully. Moreover, any overlapping resonance was naturally masked by the next sound event.

Individual samples were extracted from the long audio recordings, separating 24 audio files for each technique (6 iterations $\times$ 4 microphones). Segmentation was achieved in a two-stage process. Firstly, time points of the player's excitation onset and offset (e.g., bow movement start and stop) were marked by hand in a Praat textgrid [20], guided by audition and by viewing the waveform and spectrogram of a DPA 4060 microphone recording. Secondly, the Praat textgrid was read in Matlab [21], where time boundaries were moved to the nearest zero-crossing (independently for each microphone channel) to reduce artefacts in subsequent signal analyses. The samples were then excised individually, normalised to achieve a consistent root-mean-square level across the entire group, and rescaled *en masse* to be saved as 44.1 kHz, 16-bit WAV files without clipping.

2.4 Annotation with Timbral Descriptors

Automatic annotation was undertaken in Matlab using the Timbre Toolbox [22]. Every audio file was individually analysed to obtain an array of numerical values, each one a 'timbral descriptor' that characterises an aspect of the signal.

Instinctively, we aim to represent aspects of the audio in such a way that the performer can repeat a technique and have it interpreted in the 'same' way by the machine each time. To best match human audition, we reason that the unimportant variation across a single technique due to either recording conditions or performer iteration should be characterised by a small variance in an ideal parameter. Furthermore, that parameter should register a large change due to the important variation in sound when the instrumentalist selects a contrasting technique to perform.

An exhaustive parameter search is beyond the scope of the current pilot study. However, Peeters et al. [22] report the importance of the central tendency and temporal variability of spectro-temporal properties, the temporal energy envelope and the periodicity of the signal. We inspect variation inherent in recording condition and performer iteration according to the first and last of these, using spectral centroid and spectral flatness measures.

2.5 Measure of Variability

The variability of human repetition was measured by means of the quartile coefficient of dispersion (QCD), a relative and dimensionless measure of variation [23]. First, the inter-quartile range (iqr = Q3 - Q1) and median parameter values were derived for individual audio samples by time-varying, frame-based analysis methods in the Timbre Toolbox [22]. Secondly, QCD quantified the quartile deviation ($= iqr/2$) as a percentage of the median,

$$QCD = \frac{iqr}{2} \times \frac{100}{median}, \tag{1}$$

such that a stable parameter results in a low QCD value (close to zero). Conversely, a large QCD value implies a high degree of variability.

3 Illustrative Examples

Variation arising from alterations in recording strategy (§ 3.1) and from performance iteration (§ 3.2) are illustrated in this section with perceptually-correlated parameters addressing spectro-temporal variation and periodicity.

3.1 Recording Strategy

Figure 1 displays the spectral centroid median of the Short-Term Fourier Transform (STFT) power spectrum (squared amplitude) for a single pitch, A_3, played with standard bowing technique at six different positions on the instrument. The mean and standard error of the *median* parameter are displayed across six performance iterations at each of the four microphone positions. A closely related value, the spectral centroid *mean* has been previously correlated with the 'brightness' of a given sound [24]. The open string (string 2) could thus be said to have had a 'brighter' timbre than the stopped strings (strings 3–7), independent of the microphone used. A paired-samples t-test, with Bonferroni corrections for two comparisons, revealed an effect of microphone placement for the Neumann microphone pair, with $t_{(35)} = 16.64$ and $p < 0.001$. However, the DPA microphones produced statistically indistinguishable spectral centroid median values irrespective of the string under which they were positioned ($p = 0.16$). It appears from figure 1 that the 'close' Neumann microphone resulted in centroid measures approximating those recorded by the DPAs for the higher numbered strings. The 'far' microphone, however, resulted in higher centroid values, exhibiting colouration of the sound attributable to the room position.

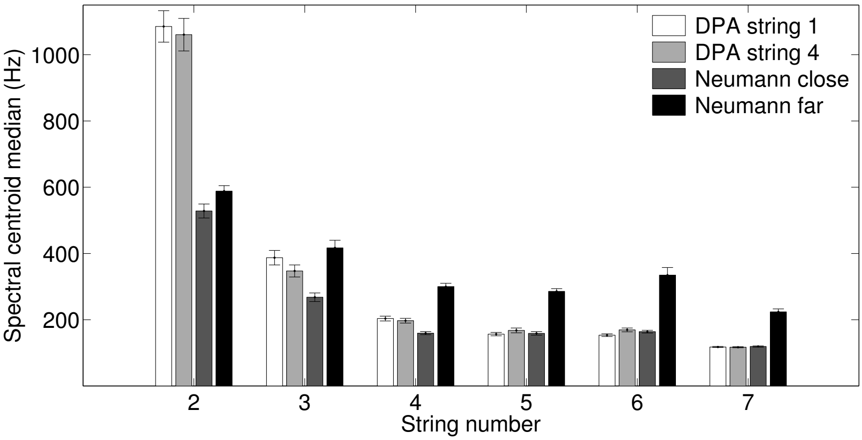

Fig. 1. Mean and standard error of six performance iterations for standard bowing of the pitch A_3 played on strings 2 to 7, as captured by the spectral centroid median of the Short-Term Fourier Transform power spectrum. The open string (2) showed a high centre of gravity, especially when recorded by the DPAs. For the stopped strings (3–7) the three 'close' microphones recorded lower values than the 'far' Neumann microphone.

3.2 Performance Iteration

The following analysis incorporates data from the three 'close' recording strategies, using the on- and off-instrument microphone positions that are frequently encountered in performance (cf. § 2.2). Figure 2 presents a range of performance techniques for the pitch A_3, ranked according to their variability, QCD, derived from the spectral flatness of the STFT power spectrum. The spectral flatness measure approaches 1 for 'noisy' timbres (with characteristically flat spectra) and 0 for 'peaky' spectra comprising sinusoidal (tonal) components [22]. Here, it is not the flatness parameter value that is of primary interest, but rather its consistency through time (in a series of consecutive frames). When the balance of noisy and tonal components alters throughout the duration of the audio file (and a larger iqr results for a given $median$ value), the QCD increases. Figure 2 clearly shows that the four extended techniques (ranking 1–4) were measured as being inherently unstable in comparison to the standard bowing techniques (ranking 5–10). Though extended techniques are often reported to be difficult for a player to reproduce consistently, the variability recorded in this data was in fact similar to that of standard bowing on the open string (the error bars are of comparable size).

4 Discussion

To better understand the reproducibility of timbral variation on an instrument, two perceptually-correlated parameters were selected from the Timbre Toolbox

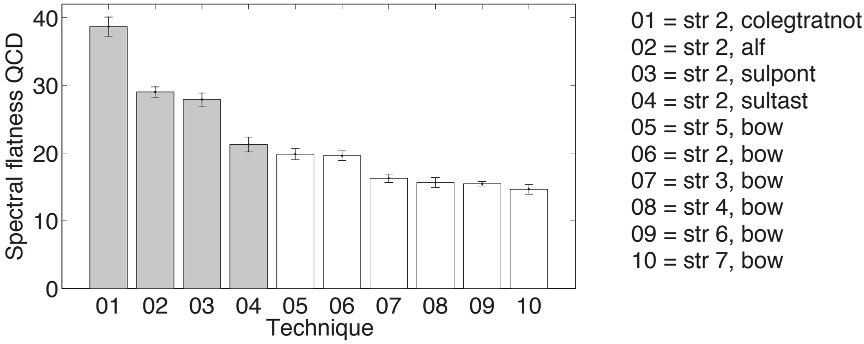

Fig. 2. Ten techniques ranked according to their quartile coefficient of dispersion (QCD) derived from the spectral flatness of the Short-Term Fourier Transform power spectrum. Extended techniques (*shaded*) were unstable and resulted in high QCD scores (*col legno tratto* – with the wood of the bow (notched from previous use); *anomalous low frequencies* – frequencies below the fundamental of the string (subharmonics); *sul ponticello* – bowing at the bridge; *sul tasto* – bowing over the fingerboard). The standard bowing techniques (*unshaded*) were more stable in their spectral flatness measure throughout the full duration of the sound, and achieved lower QCD scores.

[22] and used to quantify variation attributable to the microphone setup and to human reproducibility for a range of normal and extended performance techniques. For standard bowing techniques, close microphone positions produced similar numerical values for the spectral centroid ($median$), irrespective of the microphone type or location (§ 3.1). However, this measure did not describe the stability of this centroid over time. When the spectral flatness parameter variation was additionally tracked throughout the duration of the sound (using iqr) and incorporated in a dispersion measurement to assess iterative consistency, extended bowing techniques were found to contain more inherent variation than normal bowing (§ 3.2).

Much work has been done in recent years to extract control parameters from audio signals in live performance. Simple parameters with small computational loads were typically favoured (e.g., pitch, duration and mean amplitude [1]; pitch, amplitude, and 'roughness' [4]). At times, richer feature sets and more sophisticated approaches have been implemented (e.g., a dynamic performance state vector approach using instrument-appropriate features [2]; hidden Markov models with spectral, pitch strength and textural features [3]). However, it appears in all these works that signal variability arising from a change of room, microphone or performance iteration was considered either negligible or irrelevant.

Our current work involves recording the complete technique set with one performer, and will widen to include additional performers, violas da gamba, and acoustic spaces. The resulting dataset will allow study of the variation inherent in extended techniques, and may eventually inform improvements in machine listening algorithms to increase their reliability on stage.

References

1. Parker, M.: Searching sound with sound: Creating sonic structures from sound libraries. In: 2nd International Verband Deutscher Tonmeister (VDT) Symposium (2007)
2. Young, M.W.: NN Music: Improvising with a 'living' computer. In: International Computer Music Conference (ICMC), pp. 508–511 (2007)
3. Van Nort, D., Braasch, J., Oliveros, P.: A system for musical improvisation combining sonic gesture recognition and genetic algorithms. In: 6th Sound and Music Computing Conference (SMC), pp. 131–135 (2009)
4. Hsu, W.: Strategies for managing timbre and interaction in automatic improvisation systems. Leonardo Music J. 20, 33–39 (2010)
5. Tremblay, P. A., Schwarz, D.: Surfing the waves: live audio mosaicing of an electric bass performance as a corpus browsing interface. In: 10th International Conference on New Interfaces for Musical Expression (NIME), pp. 15–18 (2010)
6. Biber, H. I. F.: Battalia (1673)
7. Strange, P., Strange, A.: The contemporary violin: extended performance techniques. University of California Press, Berkeley (2001)
8. Turetzky, B.: The contemporary contrabass. 2nd ed. University of California Press, Berkeley (1989)
9. Kimura, M.: How to produce subharmonics on the violin. J. New Music Res. 28(2), 178–184 (1999)
10. Stilp, C. E., Alexander, J.M., Kiefte, M., Kluender, K.R.: Auditory color constancy: calibration to reliable spectral properties across nonspeech context and targets. Atten. Percept. Psycho. 72(2), 470–480 (2010)
11. Watkins, A. J.: Perceptual compensation for effects of reverberation in speech identification, J. Acoust. Soc. Am. 118(1), 249–262 (2005)
12. Fritts, L.: University of Iowa Musical Instrument Samples (MIS), `http://theremin.music.uiowa.edu/MIS.html`
13. Goto, M., Hashiguchi, H., Nishimura, T., Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database. In: 4th International Conference on Music Information Retrieval (ISMIR), pp. 229–230 (2003)
14. Opolko, F., Wapnick, J.: The McGill University Master Samples collection on DVD (3 DVDs). McGill University, Quebec, Canada (2006)
15. Vienna Symphonic Library, `http://www.vsl.co.at`
16. UVI, IRCAM Prepared Piano, `http://www.uvi.net/en/composer-tools/ircam-prepared-piano.html`
17. UVI, IRCAM Solo Instruments, `http://www.uvi.net/en/composer-tools/ircam-solo-instruments.html`
18. Eerola, T., Ferrer, R.: Instrument library (MUMS) revised. Music Percept. 25(3), 253–255 (2008)
19. Audacity, version 1.3.14, `http://audacity.sourceforge.net`
20. Praat, version 5.0.40, `http://www.praat.org`
21. Matlab, version 8.1.0.604 (R2013a), `http://www.mathworks.co.uk`
22. Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., McAdams, S.: The Timbre Toolbox: extracting audio descriptors from musical signals. J. Acoust. Soc. Am. 130(5), 2902–16 (2011)
23. Francis, A.: Business mathematics and statistics. 6th ed., pp. 150–155. Thomson Learning, London (2004)
24. Grey, J. M., Gordon, J. W.: Perceptual effects of spectral modifications on musical timbres, J. Acoust. Soc. Am. 63(5), 1493–1500 (1978)