Towards an Aesthetic of Instrumental Plausibility for Mixed Electronic Music

Richard Dudas
CREAMA
Hanyang University
Seoul, Korea
dudas@hanyang.ac.kr

Pete Furniss
Reid School of Music
Edinburgh College of Art
University of Edinburgh, UK
p.furniss@ed.ac.uk

ABSTRACT

The implementation of live audio transformations in mixed electronic music raises the issue of plausibility *in real-time instrumental transposition. The composer-performer collaboration described in this paper deals with two of the composer's existing pieces for solo instrument and computer, addressing issues of timbre and intonation in the output and adapting the existing software with improvements informed by both the physical resonating properties of musical instruments and by instrumental ensemble practice. In preparing these pieces for publication, wider performance and further instrumental transcription, improvements stemming from both compositional and performative considerations were implemented to address this issue of plausibility. While not attempting to closely simulate a human approach, the authors worked towards a pragmatic heuristic that draws on human musical nuancing in concert practice. Alternative control options for a range of concert spaces were also implemented, including the configuration of user input and output at interface level in order to manage common performance-related contingencies.*

1. INTRODUCTION

Following several years of collaboration on the performance of mixed electronic music, the authors decided to return to two existing pieces, *Prelude I for Clarinet and Computer* and *Prelude II for Clarinet and Computer*, in order to modify and update their technological component. The primary motivation behind this was simply to make the audio processing "sound better" for the purposes of including them on a published sound recording. The second motivation was to prepare the pieces for wider dissemination, including transcription of one of the pieces for a variety of other instruments, through publication of the score and technical materials. Both motivations necessitated an updating and refinement of the underlying audio processing. Improvements to the audio signal processing were geared toward the implementation of a *plausible* instrumental transposition – one that is informed by physical resonating properties of instruments and by instrumental ensemble practice. In doing this, we were not attempting to simulate a human approach, but rather create something pragmatic that draws on human musical nuancing.

The desire for a plausible instrumental transposition required addressing the way that audio effects can modify the perception of instrumental resonance in uneven ways. While resonant filter banks have been used frequently to simulate instrumental resonance where sound synthesis is concerned [1, 2], here they were employed to provide a homogeneity within the transposed material. Furthermore, the recent move to 64-bit has brought subtle improvements to clarity in audio signal processing that become musically significant in multi-layered musical and sonic textures. It was therefore necessary to make updates at the code level to some project-specific software. Finally, it was decided to break from an exclusive use of equally tempered semitones as a subtle step in an attempt to impart a chamber music aesthetic to the computer processed output.

The work undertaken on instrumental transcription and interface design represents a continuation of the authors' earlier research on *Prelude I*. [3] Improvements to the interface and audio processing chain were implemented in order to address the configuration of user input and output for the management of common contingencies in the performance space. The existing user interface was further adapted to provide a diversity of options for control either onstage by the performer or offstage by a technical assistant.

2. RESONANCE AND FORMANT FILTERING

The two pieces referred to here each employ real-time transformations of the live input, including transposition both within and beyond the actual range of the instrument(s) to create the effect of a virtual ensemble. Where these transpositions extend beyond a perfect fifth in either direction, the question of plausibility becomes an issue [4] in respect of an overall ensemble aesthetic. In the case of the *Prelude I*, transcription of the original solo flute part to both violin and viola had led the composer to adapt the software to string instruments by first filtering out the fixed formant structure of the resonating instrumental body using notch filters before transposition, and later adding the formants back into the transposed sound through the use of resonant filtering. This creates a greater sense of homogeneity, since the formants remain stable when the sound is transposed. *(Refer to Sound Example Set 1 via the link provided at the end of this paper.)*

In comparison to stringed instruments, the spectral envelope of woodwind instruments is heavily dependent on both pitch and volume (this is especially true of the clarinet and flute), and does not have an entirely fixed formant structure. However, a generalized pitch-invariant formant-based model has been shown to be helpful in improving perceptual evaluation of synthesized instrumental tones [1, 2], so the technique used with the string versions of the piece seemed potentially appropriate for any woodwind transcription. Therefore, in addition to including formant filtering in the violin and viola transcriptions of *Prelude I*, it was decided to introduce a similar filtering system to the clarinet version of the piece [3], in order to improve the instrumental perception of transposed audio material. [1] [5] Had string versions of this piece not been created, the use of fixed formant filtering would probably not have been considered for wind instrument transcriptions, however, since the filtering had been implemented in the performance software and proved to be effective with the clarinet when tested empirically, it was added to the clarinet version.

Wind instruments are inconsistent in their resonating tube lengths across the range of musical pitches available, in comparison to other instrument classes. [6, 7] Orchestral stringed instruments, the guitar and the majority of percussion, for example, maintain much the same resonating body over a variety of frequencies. In valved brass instruments, there are a small number of differing resonating tube lengths, according to which valves (or combinations thereof) are depressed, or which harmonic is being emphasised by the embouchure of the player. The trombone has a relatively consistent, though smoothly scalable resonating tube. In order to assimilate a plausible overall resonance for the clarinet, we decided to take samples from each part of the instrument's range using *glissandi* in order to find the prominent stable resonant frequencies in the spectrum, which could be filtered out before transposition and added again afterward. This is consistent with the successful implementations in the string versions of the piece: the resulting transpositions become more credible in terms of the instrumental textures they are modeling. *(Refer to Sound Example Set 2 via the link provided at the end of this paper.)*

3. UPGRADING DIGITAL AUDIO RESOLUTION

Upgrading the software to 64-bit sample compatibility was initially enacted out of the necessity to maintain performance software. [3] However, side-by-side comparisons of 32-bit and 64-bit audio output were also found to present higher clarity and definition, particularly noticeable in musically dense passages. [2] *(Refer to Sound Example Set 3 via the link provided at the end of this paper.)* On a technical front, the update of the MSP external objects themselves was fairly straightforward: it simply required making minor modifications to the code in accordance with the specifications in the most recent Max Software Developer's Kit and recompiling.

A sample rate hike to 48kHz or much higher was discussed, which could potentially further improve the quality of transposed sounds from the live input. However, although the recording industry currently leans towards 96/192kHz, such a retrofit would have required considerable recoding of the patch, which was considered overly laborious to be of notable benefit at the present time. Nevertheless, higher rates will certainly be investigated for future pieces in this series.

4. INTONATION

> *"The question of intonation is evidently relevant to nearly all instrumentalists ... and has a profound influence on the way composers and performers collectively think about harmony and intonation."* – Mieko Kanno

Where real-time transpositions are concerned, manually hard-coding intonation choices is burdensome and time-consuming, even with the aid of the computer to help pre-calculate ratios to semitones; this is especially true when compared to the immediate and multiplex tuning adjustments that trained musicians make by ear. [8, 9, 10, 11, 12, 13] Therefore it was decided to create an algorithm to automate intonation for real-time transposed chordal structures defined in the score in terms of semitones for these pieces, in lieu of simply providing transposition in cents, or fractions thereof. This is not an issue of user interaction but rather a compositional and aesthetic choice. There are prior examples of this kind of system, such as Eivind Groven's automat for adaptive just intonation [14, 15], the algorithm for which makes a note choice from a selection of fixed pitches. Many such systems are appropriate for keyboard instruments (Groven's was first implemented for organ), but the basis for these is quite different from the type of tuning that other instrumentalists or vocalists may intuitively execute in performance.

Kanno cites Fyk in relation to four distinctive types of "expressive tuning" employed by instrumental musicians (and singers): "harmonic, melodic, corrective and colouristic." Harmonic tuning relates to just intonation in relation to explicit or implicit vertical structures, while melodic intonation concerns a relative broadening and tightening of intervals based on melodic direction. Corrective intonation is "instinctive tuning" which occurs when a performer hears a discrepancy between projected and perceived pitches, while fine adjustments of timbre may be achieved by colouristic intonation choices. All of these ongoing manipulations of pitch require "the linearity of time against which to map out [their] expressive intention." [10]

Many common chords (such as triads or secundal/quartal harmonies) are relatively straightforward to tune. For example, a major triad consists of three justly tuned intervals: a perfect fifth (3:2 ratio or 1.5) defined by the outer pitches, a major third (5:4 ratio or 1.25) and a minor third (6:5 ratio or 1.2), both delineated by the central pitch?s relation to the two outer pitches. Each of the just intervals aligns perfectly with the others to form the triad (1.25 * 1.2 = 1.5). This would not be the case for a three-note vertical structure with a sharp 4th degree such as C F# G, since the just ratios for the tritone (11:8 or 1.375) and minor second

[1] It was not necessary to revisit the original flute version since the basic spectral correction used for that version was already taken into into consideration during the compositional process.

[2] For a single stream of audio processing there is often little to no perceptual difference between the two, however there *is* an audible difference when mixing multiple audio streams. This could have something to do with dither being at a lower volume when mixing multiple sources, or just an increased amplitude resolution mitigating high frequency phase cancellation when mixing multiple signals.

(17:16 or 1.0625) do not superimpose to comprise a justly tuned perfect fifth (1.375 * 1.0625 = 1.4609375 not 1.5). [3] This implies that, if we want the outer notes to delineate a justly tuned perfect fifth, we need to make a choice between tuning the F# to the C (with a just tritone), or tuning it to the G (with a just minor second), since it cannot be justly tuned to both notes and yield an in-tune perfect fifth between the outer notes. In each of scenarios the F# will cause beating with one of the two other notes. Alternatively, we could split the difference between tritone and minor second and tune the central note somewhere in-between, so it beats evenly (or even unevenly) with both of the (in tune) outer notes. For many musicians, including string players but also singers and players of wind or brass instruments, the fine adjustment of this intonation is an instinctive, internalized process based on years of experience and deliberate practice. For the computer, however, it is rather more complicated, since the programmer must create an algorithm to find the appropriate tuning nuance in each case.

The algorithm used here measures the intervallic content of chordal structures, calculates individual frequency ratios for each interval, and adjusts the calculated frequency of each note based on the weighted consonance of each interval within the chord [16, 17, 18], with respect to a reference pitch in the chord (usually the note being played by the live instrumentalist in mixed electroacoustic music). Once a midi note is identified from the input, it is converted to a frequency value and thereafter dealt with on a ratio basis instead of using semitones and cents. A very slight amount of random variation proportional to the frequency of any given pitch is added to the final tuning to emulate human error, and keep highly justly tuned chords from becoming static, seeming too mechanical or perceptually fusing into a single note. This humanizing of the intonation is heuristically modeled on the various types of tuning that string players perform for double-stops, or that the individual performers in chamber or vocal ensembles (without piano) perform when tuning chords. [10, 11] The outcome is that such a system can be used in future works, or retrospectively implemented in other compositions in which a system of flexible, unequal-tempered tuning improves the *instrumental plausibility* of the electronic effects used in that composition. As always, the objective is not to attempt to faithfully simulate a human approach, but rather create a pragmatic method inspired by human performance that draws on musical nuancing. *(Refer to Sound Example Set 4 via the link provided at the end of this paper.)*

5. INPUT AND OUTPUT

The software was restructured to allow for various user-configurable inputs, as well as gain and balance controls at various stages in the processing path. This is often overlooked, but allows the patch to be tailored to a variety of different performance scenarios. Similar one-patch-fits-all approaches have been used by a variety of composers (e.g., Kaija Saariaho, Martin Parker and Alexander Harker among others). It was also noted here that an increasing number of performers in the field are preferring to run their electronic parts directly, without the aid of an offstage technician. These user-oriented input and output controls within the onscreen interface allow for relatively rapid and simple adjustments to be made by the performer *in situ*, in response to aspects of the performance space, such as balancing microphone signals into the system, setting live direct output level and managing feedback with a variety of input solutions.

While many clarinetists employ two or more microphones to adequately cover the range of the instrument, an additional, isolated (eg. piezo or contact microphone) input may be used to track pitch without feedback from the software output. [3] Although the quality of these transducers may be inferior to a high quality condenser or ribbon microphone [4] , it was considered advantageous to provide for the use of up to three inputs for any instrument, with the ability to mix relative levels directly from the user level of the software interface. In our case, the third input was disregarded entirely for output purposes (being used only to feed the pitch tracker). This allows a performer to make adjustments quickly according to varying performance ecologies (the acoustic space, loudspeakers, microphones etc), which may differ considerably to rehearsal conditions, in order to manage feedback and projection levels. *(Refer to Sound Example Set 5 via the link provided at the end of this paper.)* A continuation of this input/output configurability relies on performers adopting strategies such as presets, VST plug-ins, or drop-ins [19], in order to assert established priorities regarding their overall sound.

The restructuring also involved separating the sound sources produced in the piece from a fixed speaker definition, so it can be performed with any given multichannel (or stereo) speaker configuration. It was previously limited to 4 or 2 channel output – it now deals with spatial location using a standard azimuth definition [5] – presuming that the speakers are arranged more or less along a circular path around the hall, when using a multichannel setup. (In a stereo scenario the left-right panning information is extracted from the azimuth.)

6. CONCLUSIONS

In implementing the above improvements to two existing pieces, it was discovered that the additional filtering before and after transposition was worth the effort in terms of a timbral strengthening of the instrumental plausibility in the electronic part. Furthermore, although the upgrade to 64-bit signal processing was dictated by the necessity of software maintenance, it was a pertinent element in improving the clarity of the electronics, particularly in densely scored sections.

Further research to determine the perceptual effectiveness of intonation adjustments could be undertaken using a music perception experiment, with the results used to determine improvements to the algorithm. In the two pieces discussed in this paper, both instrumental plausibility and

[3] The same holds true if other ratios such as 7:5, 10:7, 24:17, 45:32, etc. are used to define the tritone.

[4] Considerable improvements continue to be made in this area. For example, we use the Rumberger K1X Advanced Piezo Technology condenser microphone mounted within a specially adapted clarinet barrel. The relatively high quality of this device gives further options in the blending of input sources to accurately reflect the performer's sense of priorities in terms of their sound.

[5] Vector based panning (VBAP) [20] was used to accomplish this.

effective transcription [3] were enabled by the adaptation of filtering and an approach informed by both the physical properties of instruments and the needs and priorities of their players within a variety of performance environments.

Acknowledgments

This research was part-funded by the Arts and Humanities Research Council, UK (AHRC). The authors are also grateful to hosts Hanyang University College of Music and the University of Edinburgh (Edinburgh College of Art). Thanks also to Dr. Miriam Akkermann, for her input and assistance during stages of this research.

7. URL

http://www.richarddudas.com/ICMC2016Sounds/

8. REFERENCES

[1] S.-A. Lembke and S. McAdams, “A spectral-envelope synthesis model to study perceptual blend between wind instruments,” in *Proceedings of the Acoustics 2012 Conference*, Nantes, 2012, pp. 1031–1036.

[2] ——, “The Role of Spectral-Envelope Characteristics in Perceptual Blending of Wind-Instrument Sounds,” *Acta Acustica united with Acustica*, vol. 101, pp. 1039–1051, 2015.

[3] P. Furniss and R. Dudas, “Transcription, Adaptation and Maintenance in Live Electronic Performance with Acoustic Instruments,” in *Proceedings of the Joint International Computer Music Conference and Sound and Music Computing Conference*, Athens, 2014, pp. 456–462.

[4] S. Bernsee. Time Stretching And Pitch Shifting of Audio Signals – An Overview. [Online]. Available: http://blogs.zynaptiq.com/bernsee/time-pitch-overview/

[5] R. Dudas, “Spectral Envelope Correction for Real-Time Transposition: Proposal of a ‘Floating-Formant’ Method,” in *Proceedings of the International Computer Music Conference*, Gothenburg, 2002, pp. 126–129.

[6] A. H. Benade, *Horns, Strings, and Harmony*. Dover Publications, 1992[1960].

[7] N. Fletcher and T. Rossing, *The Physics of Musical Instruments, 2nd Edition*. Springer Verlag, 1998.

[8] K. Sassmannshaus. [Online]. Available: http://www.violinmasterclass.com/en/masterclasses/intonation

[9] R. W. Duffin, *How Equal Temperament Ruined Harmony (and Why You Should Care)*. W. W. Norton and Company, 2007.

[10] M. Kanno, “Thoughts on How to Play in Tune: Pitch and Intonation,” *Contemporary Music Review*, vol. 22, no. 1-2, pp. 35–52, 2003.

[11] S. Fischer, “Intonation,” *The Strad*, vol. 125, no. 1495, pp. 78–79, 2014.

[12] ——, *Basics*. London, UK: Peters Edition Ltd., 1997.

[13] L. Auer, *Violin Playing as I Teach It*. Toronto: Dover Publications, 1980[1921].

[14] J. Rudi, “Eivind Groven’s automat for adaptive just intonation: A pioneering example of musically situated technology,” *Studia Musicologica Norvegica*, vol. 41, pp. 40–64, 2015.

[15] ——, “The Just Intonation Automat – a Musically Adaptive Interface,” in *Proceedings of the International Computer Music Conference*, Denton, 2015, pp. 42–45.

[16] A. Plomp and J. L. Levelt, “Tonal Consonance and Critical Bandwidth,” *Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560, 1965.

[17] A. Kameoka and M. Kuriyagawa, “Consonance theory, part I: Consonance of dyads,” *Journal of the Acoustical Society of America*, vol. 45, no. 6, pp. 1452–1459, 1969.

[18] ——, “Consonance theory, part II: Consonance of complex tones and its calculation method,” *Journal of the Acoustical Society of America*, vol. 45, no. 6, pp. 1460–1469, 1969.

[19] M. Edwards. input-strip. Open source Max/MSP software. [Online]. Available: http://www.michael-edwards.org/software/

[20] V. Pulkki, “Compensating displacement of amplitude-panned virtual sources,” in *Audio Engineering Society 22th Int. Conf. on Virtual, Synthetic and Entertainment Audio*, Espoo, 2002, pp. 186–195.