Feigning Humanity: Virtual Instruments, Simulation and Performativity

Eve Klein

University of Queensland

e.klein@uq.edu.au

Abstract

This article is concerned with the ways virtual instrument software simulates acoustic human performance. In particular, it examines two case studies – virtual orchestral instruments and virtual singing instruments – to consider how their design and implementation seek to express human music performance by adopting the micro and macro sonic variations of timing, pitch, dynamics, articulation, ambience, and other limitations imposed by the physical relationship between the player and the instrument. Baudrillard considers that "simulation threatens the difference between the 'true' and the 'false', the 'real' and the 'imaginary'" (1994: 3). By feigning the acoustic markers of expressive human musical performance, virtual instrument designers and composer-users encourage the listener to produce, in themselves, the experience of hearing an orchestra or singer. Users also contribute to the recontextualization of human performance by feeding back into the cultures and development cycles of virtual instrument software, where sonic gestures are recurrently refreshed. The construction of virtual instruments as devices of musical expressivity is, therefore, an evolving, mutually constructed, and performative endeavour.

KEYWORDS: virtual instruments, simulation, performativity, virtual orchestras, musical expressivity, synthesizers, Vocaloid

Introduction

This article is concerned with the ways virtual instrument software (VI) simulates acoustic human performance. In particular, it examines two sub-categories of Vis – virtual orchestral instrument software (VOI) and virtual singing instrument software (VSI) – to consider how their design and implementation seek to express human music performance by adopting the micro and macro sonic variations of timing, pitch, dynamics, articulation, ambience, and other limitations imposed by the physical relationship between the player and the instrument. Baudrillard considers that "simulation threatens the difference between the 'true' and the 'false', the 'real' and the 'imaginary'" (1994: 3). By feigning the acoustic markers of expressive human musical performance, VI designers and composer-users encourage the listener to produce, in themselves, the experience of hearing an orchestra or singer.

This article will further consider how the simulation of orchestral music is performative within the user communities, work cultures, and audiences which embrace them. It will also examine Vocaloid virtual voices as a kind of drag act, playing out, through their sounding, the acts and gestures of human musical expressivity. Like drag, VSIs antagonize some listeners because their sounding calls into question human musical performance by displacing it. The traces of VI production – an imperfect legato, a glitch in sung text – reveal the imitation, but in doing so they also highlight the ongoing construction behind the production of such devices in human musicians, where the glitch is a variation that signifies humanity. Consequently, notions of "humanness" and "expressivity" in musical performance can be understood as being continually resignified through the repetition of particular acts and gestures which are open to recontextualization. Users contribute to this recontextualization by feeding back into the cultures and development cycles of VIs, where sonic gestures are recurrently refreshed. The construction of virtual instruments as devices of musical expressivity is, therefore, an evolving, mutually constructed, and performative endeavour.

VIs are kinds of computer software which can be played or programmed to reproduce musical sound. VIs may be used as stand-alone software applications, or as software plug-ins employed in combination with related music software such as digital audio workstations and notation software. Additionally, many VIs can be triggered by external music hardware such as keyboards and drum pad controllers. VIs are commonly played or programmed using the MIDI (Musical Instrument Digital Interface) standard which is used to describe musical performance across music hardware and software. Musicians employ VIs for a range of reasons including economy, convenience, familiarity, access to a greater variety of sounds, control of the musical performance, ease of realizing desired musical or sonic textures, or an aesthetic preference for the sounds produced (Morgan 2015; Pejrolo and DeRosa 2007: xi-xxi). Historically, VIs have been divided into two categories: synthesizers which generate sounds based upon input data, or samplers which trigger pre-recorded audio samples based upon input data (Brown 2015: 142-144). Contemporary VIs disrupt this binary approach with many plugins realizing playback through a combination of sampled, generated, and processed sounds (Klein 2015a: 15).[1] An example of this is iZotope's Iris 2 VI which the company describes as a "sample-based synthesizer … combining the

power of a sampler, the flexibility of a modular synth, and the fun of spectral filtering" (iZotope 2016: web source). Arguably, there are aesthetic distinctions between the timbral goals of VIs, with some attempting to credibly reproduce the sounds of acoustic instruments (see: East West 2016a), some seeking to creatively explore abstract sonic textures without reference to acoustic instruments (see: Native Instruments 2016), others trying to emulate the sound of popular music production hardware such as analogue synthesizers and drum machines (see: Arturia 2016), and some taking a hybrid approach (see: Spectrasonics 2016). This paper is primarily concerned with VIs which try to emulate the sounds of acoustic instruments and human musical performance, in particular VOIs and VSIs. Importantly, VOIs and VSIs may sample, synthesize and/or process sounds in complex ways in order to achieve a realistic emulation of human musical performance.

Humanness in musical performance

Historically, VIs have been received with suspicion and/or outright hostility from segments of the musical community who view them as displacement technologies (Klein 2015b; Godlovitch 1998: 69). Musicians' unions opposed these technologies because they were perceived to threaten the employment opportunities of members (Local 802 2003), while music critics considered that such technologies could have a debilitating impact on human skill, devaluing human performers and threatening them with redundancy (Godlovitch 1998: 69). Both groups sought to undermine the perceived value of VI technologies by calling into question their timbral credibility, musical expressivity and technical craft (see: Local 802 2003; Godlovitch 1998: 64-68; Clarke 1998: 90). Despite the widespread adoption of VIs in the present day such criticisms persist, particularly in relation to classical music and VOIs (see: Davies 2011: 167; Klein 2015b; Morgan 2016: 84; Player 2014; and Talk Classical 2011).

Performance, when realized by a human player, can be deemed valuable because it is expressive (Lindström et al. 2003: 23).[2] Woody, summarizing the scholarship of musical expressivity, argues that despite competing philosophies describing the relationship between music and human expression, there is consensus around the idea that music is expressive (Woody 2000: 14). Woody also notes that "most performing musicians intend to communicate some kind of meaning (emotion, feeling, or mood) in their performances, and listeners tend to hear this expressed in music" (ibid). Woody argues that communication of expressivity in music can be understood as having several sequential components encompassing "the inspiration of the composer, … the written score, the interpretive and expressive intentions of the performer(s), the produced sounding music, and the perception and emotional response of the listener" (ibid). Under this model, "expressive intentions" and the "produced sounding music" remain the responsibility of performers (ibid). However, listeners do not have access to the internal intentions of a musician, relying instead on the "perceivable properties of the musical performance" (ibid). Consequently, Woody – describing the work of Kendall and Carterette (1990) – considers that the expressive

intentions of a performer are accomplished only if they are translated or "recoded" into the acoustic properties of sound (Woody 2000: 14).

Considering how expressive intentions can be translated into acoustic properties raises the question of what specific sonic properties or devices might be considered expressive and therefore affectively human. Palmer states that it is the microstructure of a performance exhibited by "minute variations in timing, pitch, intensity, and timbre that characterize a human performance" (Palmer 1989: 331). Such microstructures may be intentional, random, automated/unconscious, or involuntary (ibid). Consequently, humanness in musical performance may be signified or "encoded" (Juslin 2000: 1798) by employing expressive acoustic devices encompassing:

–   nuanced variation in timing and accent (meter, micro and macro tempo changes, pauses/breaths, syncopation/suspension/anticipation…) (see: Palmer 1989; Juslin 2000: 1797-1799; Clarke 1995: 21);

–   nuanced variation in pitch including pitch instability (fluctuating or imprecise intonation…); pitch connection and/or sliding (portamento, legato, glissando…); pitch bending (microtonal variation…); periodic pulsation/repetition (vibrato, tremolo…); and harmonic/resonance effects (multiphonics, vocalisations through an instrument…) (see: Clarke 1995: 24; Clarke 1998: 87; Sundberg, Filipa and Himonides 2012: 391.e7);

–   nuanced variations in dynamics/loudness which may remain stable or change gradually or suddenly (crescendo, decrescendo, sforzando…) (see: Clarke 1995: 24; Woody 1999);

–   the nuanced application of other instrument-specific playing techniques or articulations which influence the timbre of sound production (tonguing, bowing, plucking, striking, breath, or muting techniques) (see: Clarke 1998; Woody 2000, 14);

–   limitations on all of the above devices based upon physical interaction between the player and the instrument (the pitches which can be reached within a human hand-span; the speed and number of keys which can be depressed by a human hand at any one point in time…) (see: Clarke 1995: 27; Godlovitch 1998: 69; Clarke 1998: 90);

–   responsiveness of the performed sound to an acoustic environment (the instrument sounds like it is being played in a real-world space through the presence of sonic reflections, dampening, distortion… and that this relationship between environment and performed sound is complementary) (see: Pätynen 2016: 1223).

These musical devices impart expressivity by colouring and shaping the performance of individual notes, musical phrases and/or whole musical sections. Importantly, musicians use these devices in complex ways (Clarke 1995: 21), varying and combining devices in order to communicate musical and extramusical ideas. Such definitions of musical expressivity privilege the human listener as the site where value, via affect, is determined, with Evans stating that hearing:

> the contrast and connection between one melodic phrase and another requires a human ear and a familiarity with the musical tradition of our culture. Likewise, all the factors that index elements of form – mood,

intensity, tempo, density, verse versus chorus, etc. – rely on a trained hearing,
or at least an interested ear. (Evans 2005: 31,2)

The success of musical expression is evaluated by humans interpreting sounds within cultural and social networks where a semiotics of listening informs reception (Clarke 1995: 21). These cultural and social networks encompass modes of listening enabled by recorded media with Frith and Katz and noting the creative and performative nature of listening enabled by sound recordings (Frith 1996: 203-204; Katz 2004: 46-74).

Virtual orchestral instruments and the simulation of human performance

Recalling Woody's statement that the intentions of a performer are accomplished only if they are translated or "recoded" into the acoustic properties of sound (2000: 14), and noting that there are specific sonic devices which listeners' interpret as being humanly expressive, then such sonified human expression can also be captured via audio recording. If such expressions are entirely representable via recordings then they are, at least hypothetically, susceptible to simulation in that medium. This section will explore the strategies VI developers employ to create products capable of realizing expressive music via affective simulation of orchestral music for the listener on recorded media. How this occurs will be established by analysing the design, functionality and marketing of VOI software East West/Quantum Leap Hollywood Orchestra (EWHO), East West/Quantum Leap Symphonic Orchestra (EWSO) and Vienna Symphonic Library (VSL).

Present generation VOIs typically employ a combination of sampling, effecting, processing, and other means of sound manipulation for optimizing realistic playback within the VOI interface. Software companies create high quality recordings of professional musicians performing individual notes, scales, or musical phrases as the basis of their VOI samples. These samples are then mapped to the MIDI communication protocol so that they can be triggered and manipulated by compatible music hardware and software. More advanced VOIs will feature samples of different playing techniques/articulations such as legato, pizzicato, or marcato which can alternate during playback. VOIs may also have samples recorded so that different dynamic ranges for each articulation – such as pianissimo, piano, mezzo forte, fortissimo, etc. – can be reproduced. Samples are carefully edited, then packaged into a user interface, usually a MIDI-controlled software plugin for use in digital audio workstations (DAW), or as a sound library embedded in a piece of music hardware such as an electronic keyboard or synthesizer. Importantly, the playback of samples is only one component of most contemporary VOIs, with instruments such as VSL and EWHO enhancing sample playback in complex ways enabled by the playback software.

VOIs typically have an interface allowing deep performance control which can either be triggered during real-time playback or programmed using MIDI protocols and DAW-based mix automation. Sounds are organized into a hierarchy to aid user navigation (see: FIGURE 1) with the term *library* referring to the overall

instrument (for instance, solo violin), the *sound bank* referring to subcategories within that instrument (mutes, sustains, short attacks, repetitions, scales…), and *patches* referencing performance articulations within that subcategory (col legno, tremolo, spiccato…). Some VOIs also incorporate a range of features to emulate room acoustics (convolution reverb…), to modify instrumental timbre (equalizers and filters), to alter timing (for instance, quantisation), to alter pitch (transposition effects and humanize intonation controls), or to effect the outgoing signal through repetition, distortion or transformation of the audio signal (modulation effects, arpeggiators…). VOIs, particularly those aimed at film and video game composition, may include instrument patches which combine synthesis with sampled sounds to create complex patch options with a more abstracted/less naturalistic sound design (see: Spitfire Audio 2016). Many libraries also include pre-set loops, scales or chord patches to aid composition.

Musical expressivity features in the rhetoric used by software companies to promote VOIs. Many products compete in this space, including Garritan Instant Orchestra, IK Multimedia Miroslav Philharmonik 2, UVI Orchestral Suite, and Spitfire Audio Albion One.[3] Among the largest and most established products are EWSO, EWHO and VSL's product line of the same name. East West promotes EWHO as the "holy grail" of VOIs, noting the product's "unprecedented detail, superior true legato, sound quality, and sound control with five user-controllable mic positions" (East West 2016b: web source). VSL promote their competing products as "inspired by passion" and detail the creation of their sample libraries as evidence of this declaration:

> The Vienna Symphonic Library is a research-driven music software and sample library developer based in Vienna, Austria … Vienna's finest musicians invest their life-blood into recording samples that incorporate their passion for music and their instrument mastery. Likewise, our software developers strive to improve the ease-of-use of our tools and the authenticity and musicality of the results. Music creation shouldn't be limited by technology or influenced by the computer know-how of the user. Technology should enhance the flow of musical ideas and simplify the realization … Our goal has always been to create the most advanced virtual instruments to help composers, orchestrators and music producers realize their visions, get their ideas across and make themselves heard. (Vienna Symphonic Library 2016a: web source)

The above statement compacts many claims towards a human musical experience enabled by the product. Particularly noteworthy are descriptors such as "research driven", musicians investing "their life-blood" and "mastery", "passion", "authenticity and musicality", enhanced "flow", and composers making "themselves heard" (Vienna Symphonic Library 2016a: web source). Such descriptors reinforce the contribution not just of the musicians, but also of the software developers who program the VOI. In doing so, VSL aims to associate the composite sound, produced between the sampled and programmed audio, with human passion, authenticity, flow, and vision, and ultimately promise users an experience of being heard. By promoting their products in this way, VI software companies warrant that their users can access musical expressivity realized through the VOI.
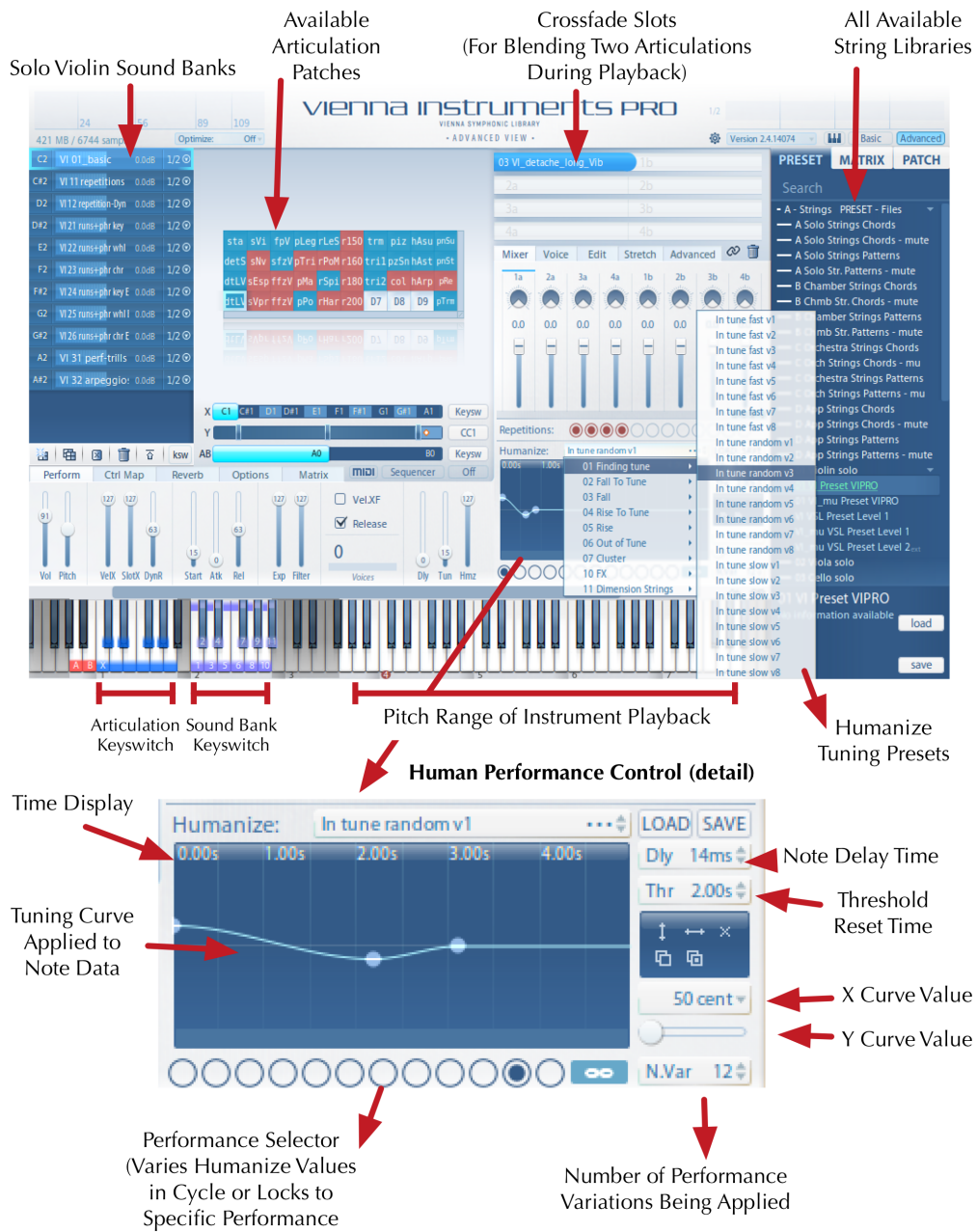
Detail of VSL Solo Violin VI

FIGURE 1: Screenshot of Vienna Instruments Pro.[4]

East West and VSL have strived to emulate the sonic devices identified, as identified in the previous section of this article, within their instrument design. Importantly, both companies have released orchestral libraries replicating solo, small ensemble, and large ensemble sonorities of all standard (and many specialized) orchestral instruments. For example, EWSO strings library includes sounds banks for three cellos, four violins, nine double basses, ten cellos, ten

violas, eleven violins, eighteen violins, harp, harpsichord, large string ensemble, and solo cello, contrabass, viola and violin. Libraries are normally organized into collections which can be extensive. For example, the flagship product for VSL is the VSL Super Package which contains all twenty-seven VSL instrument collections and is made from 2,299,916 individual audio samples, requiring 900.5 GB of hard disc space to store (Vienna Symphonic Library 2016f). Within these libraries are complex sound banks that organize musical material into instruments and/or performance styles (for instance, muted vs unmuted playback), and then into patches representing different articulations within that larger category (see FIGURE 1).

Both East West and VSL use patches to provide different articulations which can be swapped between during real-time playback to emulate the ways which human musicians move between performance techniques while playing. A patch can be simple or complex, but typically provides seamless playback for a single articulation or group of articulations. For example, the VSL Solo Violin Articulation "04 VI_detache_long_noVib" plays back samples of a performer playing long détaché notes without vibrato from G3-A#7. Samples are swapped depending upon the note velocity (attack speed) with the patch containing samples for four different velocity layers: 0–55 (pianissimo), 56–88 (mezzo piano), 89–108 (forte), and 109–127 (fortissimo) (Vienna Symphonic Library 2015b: 16). If a note is repeated with the same velocity, there are four samples which the VI will alternate between so that repetitions of a note do not sound as if they were played in the same way (Vienna Symphonic Library 2015b: 16).

EWSO Solo Violin includes thirty-eight different articulation patches (East West 2016c), EWHO Solo Violin includes thirty-four different articulation patches (East West 2015: 42-43), while VSL's Solo Violin encompasses 705 articulation patches.[5] Effectively this means that a user can program or play a musical line which jumps between legato playing, staccato playing, reproduce portamento between notes, deploy semitone or whole tone trills, and then revert back to legato playing as the score dictates. With 705 different articulation patches, VSL allows composers to reproduce most available playing techniques on the solo violin, and many playing techniques that amateur violinists would struggle to reproduce, such as extended pitch ranges and col legno. How credibly they are realized can depend upon the detail the user puts into programing these articulation changes, and related performance directions such as attack velocity, dynamic changes, tempo changes, changes in vibrato strength…

Beyond varying the playing techniques of individual notes, VOIs can emulate micro pitch and timing variations for each note played, a key expressive quality identified by music psychologists such as Palmer (1989), Juslin (2000), and Clarke (1995). A sophisticated example of this is a feature called Human Performance Control (HPC) which is included in VSL's Vienna Instruments Pro sample player. The HPC interface (see FIGURE 1) processes the playback of individual note samples by allowing the user to vary the onset of a note by manipulating: the intonation of the attack; the time it takes the intonation to settle into the central pitch; and pre-delay of the note onset. Additionally, up to twelve different variations can be applied to a MIDI track so that the performance of each note varies slightly in intonation and timing in a cyclic way. An extensive library of HPC pre-sets means that users can select whether the VOI: approaches the pitch

from below or above; wavers both above and below the central pitch; or whether the central tone is slightly sharp or slightly flat. The user can click in the Tone Curve widow to modify intonation or add additional bends in the pitch.

VSL describes the HPC feature saying it

> creates something wonderful: Human Imperfection. From subtle changes in intonation to simulating stress-induced tuning corrections and slightly un-tight rhythmic passages: Random patterns produce authenticity, be it in the string, woodwind or brass section. Additionally to these fixed variations, you can influence the intensity of the tuning curves and delay settings to create an even more random effect with real-time controllers. (Vienna Symphonic Library 2015: 49)

Describing "Human Imperfection" as "something wonderful" draws on the notion that these micro-variations in tuning and timing are what "produce[s] authenticity" in the performed sound. This evidences the company's deliberate approach to emulating micro pitch and timing devices associated with expressive human musical performance via the application of processing to sample playback.

Human performance limitations are often incorporated into VOI design. Both EWSO and VSL only reproduce notes within the playing range of professional instrumentalists and focus on replicating real-world playing techniques. They largely do not process pre-set sounds in ways which would sound artificial, but rather strive to represent orchestral instruments transparently in line with so-called "concert hall" aesthetic conventions employed in classical music recording practices (see Chanan 1995: 146; Klein 2014). Similarly, both instruments try to locate the VI performances in sympathetic acoustics, offering built-in reverberation. EWSO and EWHO incorporate convolution reverb effects with concert hall and recording studio reverb pre-sets into the VI interface. VSL includes basic reverb effects in their standard VOI player, but also sell specialized convolution reverberation effect software, Vienna MIR PRO, which provides surround sound mix placement modelled on real world acoustic environments such as the Vienna Konzerthaus and The Sage Gateshead concert halls (Vienna Symphonic Library 2016d). East West have also sought to capture room ambience when they record their instrument samples. EWHO offers up to five separate close, accent and ambient microphone positions which the user can mix together within the VOI interface to create a blend of ambience (see FIGURE 2). East West describes their multi-micing solution as allowing the user to "achieve control over both acoustic vantage and spaciousness of the sound" (East West 2015: 11). VSL describe their MIR Pro player as "allowing the user to interact with the players more like a conductor than an engineer" (Vienna Symphonic Library 2016d: web source). Though these companies have taken different approaches to the reproduction of ambiance, they readily acknowledge the importance of ambience to the reception of their VOI sound. Izhaki describes reverb as essential to the production of expressive sound recordings considering it an important device for adding character, cohesion, and vigour to a mix (2009: 409). Izhaki also conceives of reverb as "an important factor in natural simulations" which contributes to the realism of a mix (ibid). By designing ways of manipulating

ambience into VOI interfaces, software companies go beyond merely reproducing the sonic characteristics of different playing techniques by taking care to locate the listener, as someone familiar with recording aesthetics and real-world performance acoustics, in spaces such as the concert hall.

FIGURE 2: Variable Microphone Positions in EWHO.[6]

The role of MIDI is to describe musical performance in a standardized way so that it can be reproduced across electronic musical devices and software (Pejrolo 2011: 4-8). VOIs produce sound based upon received MIDI data, which describe what notes should be played, how they should be played, and for how long. Tempo adjustments of musical phrases or whole music sections can be played or programmed into a DAW or sequencer which controls the MIDI playback within an arrangement. In particular, tempo changes controlled by mix automation, such as the Global Tempo Track in Logic Pro X, allow precise and fluid tempo changes or more radical tempo jumps to correlate with specific musical events within an arrangement. Composers and producers working with video media use automated tempo changes to precisely correlate the timing of musical events to visual events (see FIGURE 3). Most DAWs possess quantisation features which will allow the user to adjust the onset (and occasionally offset) timing of notes to: align with the subdivisions of a bar; add in timing variation to a note shifting them off the subdivisions of a bar; or to vary the groove towards a swing or triplet feel. This manipulates the timing of notes so that the user can emulate, remove, or alter human feel in a given part. Finally, DAW-based automation can be used to program changes to any parameter within a VOI interface such as reverberation, articulations, intonation and pitch sliding, note onset delays... By employing mix automation the user is able to replicate the features of human playing,

programming the VOI to respond differently across time and emulating the kinds of complex changes that occur in real-world performance.
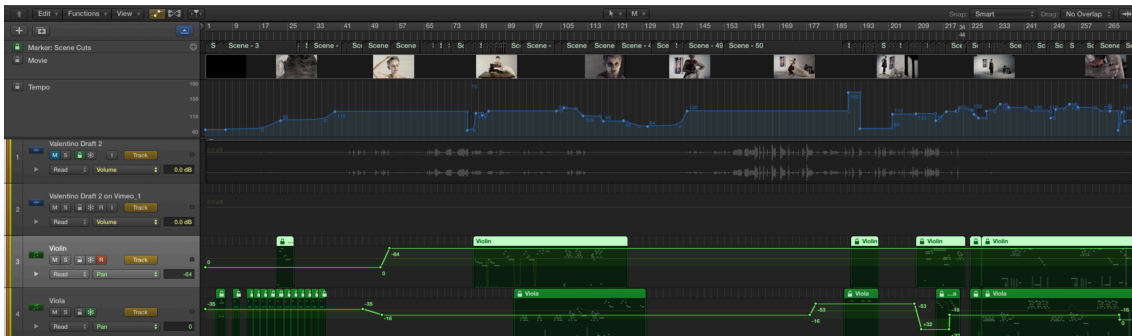

FIGURE 3: Programmed tempo changes in Logic Pro X.

Brown suggests that there are three metaphorical perspectives which inform the use of music technologies, these being the computer as a musical tool, medium and instrument respectively (Brown 2015: 6). Unpacking these ideas further, Brown considers that as a tool, music technologies provide "leverage and extension" to existing skills and human capacities, which can then be controlled by the user (Brown 2015: 6-7). As a medium, Brown considers music technologies as a vehicle for transmission with computing technologies representing, translating, manipulating, storing and transmitting musical and sonic data, with the musician acting as a kind of explorer of digital media landscapes (Brown 2015: 7-10). Finally, when considered as a musical instrument, music technologies become an active means of musical expression either through live performance or non-real-time contexts where technologies amplify musical ideas (Brown 2015: 10-11). In describing music technologies as instruments, Brown is highlighting a kind of deep engagement or relationship where the user develops "understanding and fluency" (Brown 2015: 12).

When orchestral composition is realized through detailed VOI programming in a naturalistic way, it is referred to as "MIDI orchestration".[7] Gilreth describes MIDI orchestration as "the total process of employing MIDI, samples and samplers, sound modules, processing hardware and software and recording gear to achieve maximum realism, ultimately creating a wonderful experience and sound of having a true, living orchestra within your own working studio" (Gilreth 2004: v-vii). Gilreth considers the successful MIDI orchestrator as possessing "deep knowledge of orchestration as well as the comfort and knowledge necessary to work with computers, samples and MIDI … perfecting this skill with so much realism and liveness that even the most adept musical listener will find [it] hard to tell your recordings weren't made by a real orchestra" (Gilreth 2004: v-vii). Gilreth joins a number of authors including Pejrolo (2011), Pejrolo and DeRosa (2007), and Bennett (2009), who provide detailed instructions in orchestral composition realized through VOIs with a high degree of realism. The application of music technologies as instruments becomes a pedagogical practice, the goal of which is to train the composer to produce music in a DAW or sequencer which is convincing enough that the listener, receiving its transmission,

is able believe it is being realized by a real orchestra (Gilreth 2004: vii). The assumption embedded in such approaches is the aim of producing digital music convincingly enough that the listener can engage with the underlying composition, without being unsettled by a non-expressive rendering of its "performance". In working with VIs in this way, the composer also becomes an expert software user and music producer. The casual interchange of terminology between "composer", "user" and "producer" can be seen through discussions of VOI software – and in the language employed in writing this article – pointing to broader changes in the role of composition within mediatized music cultures.

Baudrillard conceptualizes simulation as a process of "feigning what one doesn't have" (Baudrillard 1994: 3). However, Baudrillard is careful to note that simulation is not a simple absence, but rather, he likens the concept to simulating illness, stating that: "whoever fakes an illness can simply stay in bed and make everyone believe he is ill. Whoever simulates an illness produces in himself some of the symptoms" (ibid). Consequently, Baudrillard argues that pretending "leaves the principle of reality intact … whereas simulation threatens the difference between the 'true' and the 'false', the 'real' and the 'imaginary'" (ibid). According to Baudrillard this in-between state exists because our hypothetical patient could be conceived as being both ill and not ill, since "if any symptom can be 'produced'" it is no longer a "fact of nature" (ibid). Simulation therefore has a destabilizing effect because "every illness can be considered as simulateable and simulated", thereby calling into question what is real (ibid).

MIDI orchestrators, enabled by specialist VOIs which emulate human playing techniques and related nuances of timing and intonation, simulate for the listener an experience of hearing an orchestra. They do this by feigning enough of the characteristics of expressive orchestral music that the listener produces in themselves the experience of hearing an orchestra. This is similar to established practice in classical music record production whereby recording engineers seek to present heavily manipulated studio recordings as if they were a seamless performance recorded in a concert hall (see: Klein 2014) with the heard "performance" being "a joint creation between the engineer and the listener" who tricks themself "into thinking that its musical representation is real" (Symes 2004: 62). To establish this argument, let us review the characteristics of expressive music discussed previously in line with the VOI features and related music production practices discussed above. The first quality identified was nuanced variation in timing and accent. This can be achieved through both VOI-based timing manipulation such as HPC, and by DAW-based tempo automation and quantisation. The second quality identified was nuanced variation in pitch including pitch instability, pitch connection and/or sliding, pitch bending, periodic pulsation/repetition and harmonic/resonance effects. These pitch devices can be achieved largely through the selection of appropriate articulation patches, with additional pitch variation added through processing features such as HPC or MIDI based programming such as modulation and pitch bend. The third quality identified was nuanced variation in dynamics/loudness. This can be achieved by a combination of MIDI velocity and volume programming, articulation patch selections, and additional DAW mixing where appropriate. The fourth expressive quality identified was the nuanced application of other instrument specific playing techniques or articulations. This can be realized by selecting and switching

between articulations during real-time playback using either mix automation, MIDI keyswitching,[8] or MIDI channel message programming, depending upon the implementation within a VOI. The fifth quality identified was limitation of all the above devices based upon physical interaction between the player and the instrument. Makers of orchestral VOIs like VSL and East West have largely imposed such limitations in the design of their instruments, by sampling pitches and techniques from professional players and maintaining such restrictions within the playback interface. The final expressive quality is responsiveness of the performed sound to an acoustic environment. VI makers have offered a diverse range of solutions including incorporating reverb effects into their instruments, offering surround sound mixing solutions, producing convolution reverb libraries replicating the acoustics of real world spaces, and offering ambient and surround sound versions of their samples which can be mixed together according to the user's taste. Composers, musicians and producers skilled at MIDI orchestration use their VOIs, DAWs, and related equipment in complex ways combining and varying the micro and macro gestures that humans equate with expressive musical performance over time, in order to create the most realistic sounding orchestral experience for their listeners. While VOIs may be distinguishable from their acoustic counterparts for some listeners, their widespread adoption lends credence to the argument that they are successful in simulating the heard experience of orchestral music for a significant number of listeners.

Recalling Baudrillard's statement that "simulation threatens the difference between the 'true' and the 'false', the 'real' and the 'imaginary'" (1994: 3), VOIs can be seen as destabilizing acoustic orchestral music by producing enough of the qualities of orchestral music that they threaten the distinction between acoustic and virtual music creation on recorded media. By sampling, processing, effecting, and otherwise producing the micro gestures of expressive performance through the application of VOIs, composers develop syntaxes from these gestures which symbolize human orchestral playing. Such syntaxes are slowly becoming formalized into pedagogical discourses of composition such as 'MIDI orchestration' and demonstrate that it is possible to reproduce expressive renderings of orchestral music from fragments rendered in VOIs and related applications such as DAWs. This formalisation has developed to such a degree that Pejrolo considers VOIs as the "orchestras of the twenty-first century", describing the combination of orchestration and MIDI sequencing skills as essential professional practice for contemporary composers (Pejrolo 2011: 189). Pejrolo and DeRosa give the following account of why such skills are necessary:

> The consumer has come to expect a 'finished' demo that gives the truest representation of the final product…. Prior to the 1980s the composer might have played at the piano a minimal representation of the larger scale work. It took great imagination and trust on the part of the client and great inspiration on the part of the composer to convince the consumer that the endeavour was worthy of the money about to be invested. (Pejrolo and DeRosa 2007: xi)

Composers employ VOIs as a substitute for acoustic orchestras, and their clients and audiences have been convinced that their application is good enough to warrant continued use. In doing so, the primacy of the acoustic orchestra has

been destabilized. This claim is borne out by research quantifying the use of VIs by screen composers. Morgan's survey of 137 Australian screen composers found that 29% of musical cues produced by screen composers[9] were realized using VIs only, 55% of cues were realized using a combination of VIs and real instrument recordings, and only 16% of cues were realized exclusively using recordings of performers (Morgan 2014), meaning that 84% of composed screen music was realized using VIs to some degree. Morgan is careful to note that 40% of screen composers reported that they used no digital notation software in their work, only DAWs, a result which he considers indicates "that the final music delivered is generated by virtual instruments" because there is no musical score produced for human players to perform from (Morgan 2014: web source). Morgan's research also shows that of the screen composers who work exclusively or partially with VIs, the majority will avoid certain products because of their sonic properties (ibid). Respondents gave various reasons, noting that: "I try to avoid any semblance of inauthenticity when it comes to samples" and "there are certain instruments that have not quite made it to acceptable sample quality yet" (ibid). This indicates that screen composers are selecting instruments for their ability to realistically emulate orchestral sounds. Similarly, screen composers also consider the impact of VIs on their output with one respondent noting that "I think sometimes my writing might be influenced by how things sound when played back" (ibid). Morgan's follow up interviews reveal that: composers are relying upon templates in order to meet the demands of deadlines; clients do not want to hear draft versions, rather final versions of music only; composers recognize the sound of commonly utilized VIs; and that MIDI enables composers to realize music beyond their capacity to play (Morgan 2015). Morgan's research points to an industry-wide practice, whereby literacy in the use and application of VIs is both expected and enabling for composers who can satisfy clients within an increasingly time- and budget-sensitive work culture. Hollywood screen composers J.J. Abrams, James Newton Howard, Danny Elfman, John Powell, Brian Tyler and Jeff Beal all endorse East West products and speak publicly about their use in the film soundtracks they compose (East West 2016d), indicating that this practice extends well beyond Australian cultures of screen composition.

The overwhelming use of VIs in screen composition show that VOIs are successfully taking over the function of representing orchestral music in certain locations (recorded media) and forms (for instance, film and video game soundtracks), which in turn requires that composers develop new musical skills employing such technologies. A recent study (Kopiez et al. 2016) asked listeners to distinguish between recordings of an orchestra and recordings of VOIs performing Stravinsky's *The Rite of Spring*. The study found that non-expert listeners "who resembled the majority of music consumers, only achieved 68.6% … and were virtually unable to tell the real-life and OSL (orchestral sound library) sounds apart" thereby "satisfying Turing's criterion threshold of 70%" for "a convincing simulated performance" (Kopiez et al. 2016: web source). VOIs turn human musical expression into a set of discrete symbols that can be reproduced on command. If expressive human orchestral performance can be captured as a sample in micro-fashion, teased apart and recombined in the creation of digital music, then the musicians who are sampled, VOI software manufacturers, MIDI orchestrators and listeners who consume the resulting music are all participants in

the simulation of orchestral music on this media. Baudrillard says that simulation produces an in-between state because "if any symptom can be 'produced'" it is no longer a "fact of nature" (Baudrillard 1994: 3). By simulating orchestral music successfully at the level of recordings, and having those recordings disseminated across media such as film, VOIs question the aesthetic dominance of acoustic orchestral music precisely because their implementation is capable of producing orchestral music which is *expressive enough* or *indistinguishable enough* to satisfy viewing audiences, film and TV producers, and composers *themselves*.

Virtual singers, personas and performativity

The expressivity of the human singing voice represents a significant challenge for VI developers for several reasons: the demands of designing an interface capable of programming sung-text; the range of possible vocal styles and inflections which must be rendered; and the idiosyncratic nature of individual human voices as an important expressive device within musical performance. This section will unpack VSI (virtual singing instrument software) and notions of performativity to determine how voices, as a sub-category of VIs, confront the particular difficulty of rendering an instrument which originates within the human body. It will do so by discussing two separate types of VIs: those that seek to realistically simulate sung human performance, and those which seek to fabricate notions of humanness and re-contextualize it within singing personas. In particular, this section will unpack virtual choirs and Vocaloid software respectively to show how stylistic orientation impacts upon the function of expressiveness within these interfaces for human composer-users and fan communities in which this music circulates.

There are hundreds of VSIs in the market representing vocal styles as diverse as heavy metal (Soundiron's Voices of Rage), opera (Zero-G's Prima 2 and Tonio), and traditional musics (East West's Voices of Passion, Best Service's Kwaya and Soundiron's Voice of Gaia). Like other genres of VIs, some VSIs attempt to transparently reproduce the sounds of human singers, while others seek to creatively explore effected or more abstract sonic vocal timbres. Choral and classically-oriented solo vocal instruments fall into the former category. Many of the pop-oriented libraries can be classified in the latter category because they reproduce heavily effected Auto-Tune style of vocal timbres. However, the VSI interfaces driving both approaches to instrument design can be equally complex. For example, the Vocaloid VSI platform and East West's Symphonic Choirs (EWSC) both allow for the detailed programming of sung text. In the context of Vocaloid, Bell notes that the complex sampling database is "one of the keys to both the realism and the mutability of voices" (Bell 2016: 226) but that is also true of many other VSIs which seek to render the voice with reasonable detail.

The construction of musical expressivity in VSIs is comparable to VOIs in that they replicate the microstructural acoustic devices characteristic of human playing. However, because human singing is so intricately linked to the performance of text, the capacity of VSIs to realize text adds an increased layer of complexity to VI design. Developers must negotiate convincing transitions between vowels and consonants for text intelligibility, and sung text is usually

constructed phoneme by phoneme, mapped to performance articulations which can vary over time. VSIs aimed at classical music reproduction aim to simulate human vocal performances in the same ways that VOIs do, however, developers make considered choices over what is feasible within their interface, concealing where possible glitches in rendering with reverb effects and carefully programmed legato between notes. Although VSL have two sophisticated vocal libraries, Vienna Solo Voices and Vienna Choir, neither collection includes the ability to program text. The developers opted for a more limited range of vocal sounds they could credibly reproduce, which for Vienna Solo Voices encompassed:

> staccatos, sustains and Legato Performances of the vowels "A" (aah) and "U" (ooh) in four dynamic levels as well as various staccato repetitions (of the syllables "ta", "pa", "ra", "sa", and "tu", "pu", "ru", "su"), sforzatos, crescendos, diminuendos, trills and glissandos. (Vienna Symphonic Library 2016e)

EWSC instrument does reproduce sung text, but only in its choral sound banks for Sopranos, Altos, Tenors, Basses and Boys choir sections (East West 2013: 14). EWSC also includes sounds banks for Soprano, Alto, and Boy soloists, but perhaps because solo voices are more exposed, they have opted not to include text programmability into these patches. To enable the reproduction of sung text, East West designed an interface called Wordbuilder (see FIGURE 4) which gives the user control over how long each consonant and vowel are held within a syllable, including the ability to precisely mix the dynamic level of each phonic component (East West 2013: 15). Designed primarily for programming English language text, Wordbuilder contains a 100,000-word dictionary which translates American English spelling to a phonetic spelling, and the company has also created a phonetic alphabet called Votox to aid programming within the interface (East West 2013: 15).

The Wordbuilder software requires a detailed level of manipulation to produce realistic sung text. In order to realize expressive results, Wordbuilder must be programed in combination with the main software player which replicates performance articulations and related devices. For example, EWSC includes 210 patches encompassing a wide range of articulations including vibrato singing, straight-tone singing, non-pitched tones, shouts, whispers, falling and rising phrases, individual consonants and vowels, and transitions between vowels.[10] The interface also allows for the programing of portamento, legato, and variations on repeated notes. Like EWSO the user can create customised ambience by blending together close, accent, and ambient microphone positions during real time playback. All such parameters can be controlled over time with the application of DAW-based mix automation, so it is possible for the user to create intricate variations in playback expression which change across the length of the performance. East West has created extensive documentation advising how users can more effectively employ Wordbuilder, and warn the user that its effectiveness "depends on your craft in shaping a series of phonetic material into realistic patterns that the listener perceives as language" (East West 2013: 29). Given the intricacy of realizing a convincing rendition of sung text, MIDI orchestrators are likely the target user because they have already developed experience with

professional-level MIDI programming, and it is they who would be most likely to learn one of the two possible phonetic alphabets to get the most reliable results from the interface. By creating such a complex interface, the developers create a barrier which potentially filters out novice users, and selects for users most capable of realizing expressive renderings within the software.

FIGURE 4: East West's Wordbuilder Text Programming Interface.[11]

Yamaha promote their singing voice synthesis software, Vocaloid 4: Cyber Diva, with a tag line: "human beings are still singing? We are in the digital age! … It's time for you to join us!" (Yamaha 2016). The Vocaloid advertisement features close shots of a human hand controlling a DAW, playing a guitar, and tapping a drum pad, while programing the Vocaloid instrument to sing "I love you… you love me", implying an intimacy between the human producer and the virtual female singer (ibid). Vocaloid, the platform associated with the first significant range of VSIs, was released in 2004 with three different voice libraries: Leon, a male soul singer; Lola, a female soul singer; and Miriam, modelled on the voice of pop singer Miriam Stockley (Walden 2004a; Walden 2004b). While the instruments initially received a subdued reception in western markets, Crypton Future Media (Crypton) President Hiroyuki Itō saw the potential to "reinterpret" the use of Vocaloid technology for the Japanese *dōjinshi* self-publishing media

culture, releasing a Vocaloid singer named Meiko later in 2004 (Bell 2016: 227). Bell notes that "users responded, and, as the software matured, so did Crypton's strategy for selling it", evolving from a female anime drawing on Meiko's cover to the release of Hatsune Miku, a virtual singing persona released for the Vocaloid 2 platform in 2007 (ibid). Miku was given the look of an anime character but was further promoted as a sixteen-year-old girl who was 158 cm tall, weighing forty-two kilograms and whose favourite music genres were J-Pops and Dance-Pops (Crypton 2016). Crypton promote Miku's trajectory from a "vocal synthesizer product to beloved collaboratively constructed cyber celebrity with a growing user community" (ibid). They claim that Miku has been used in the release of 100,000 songs, 170,000 YouTube videos, and an 1,000,000 created artworks, that Miku has 900,000 Facebook fans, and performs sold-out 3D concerts worldwide (ibid). To date, there are 169 commercial Vocaloid voice libraries developed for nine iterations of the platform, many of whom have a basic profile or more developed personas attached.[12] Sometimes the identity of the human singer who was sampled for the instruments are also attached to the VSI, as was the case with Miriam and subsequent Vocaloid singers such as Hiyama Kiyoteru (sung by Kiyoshi Hiyama) and Luo Tianyi (sung by Shan Xin) (see Ah-Software 2016; Sh-henian 2016). By attaching real singers' identities to the product and by constructing personas for Vocaloid sample libraries, the companies developing products for the Vocaloid platform negotiate a liminal space "between the 'true' and the 'false', the 'real' and the 'imaginary'" (Baudrillard 1994:3) that, if the Cyber Diva advertising and fan culture surrounding Hatsune Miku is any indication, developers hope will further displace human singers.

Vocaloid 4 Editor, the latest version of the Vocaloid host platform, enables users to compose, edit, and mix multi-track vocal parts from Vocaloid singer libraries. The platform contains a piano roll where the primary vocal programming is contained. To make the Vocaloid singer perform text the user clicks on a MIDI note and types in lyrics which are then distributed across available notes in the sequence. Further editing control is possible by editing the phonetics of each note and the Note Properties such as vibrato. In comparison to EWSC the sung text editing component of the interface is underdeveloped, with users being able to edit individual phonemes but not control the mix between elements within a phoneme (for instance, the blend between vowels within a diphthong), instead relying upon the embedded language dictionary as the primary means of shaping the text performance. Users split MIDI notes to add in extra vowels, consonants, or breath sounds where a more fine-tuned text performance is required. However, Vocaloid have enabled substantial control over the performance style of individual notes with parameters for qualities like "growl", "pitch bends", "gender factor", "breathiness", "brightness", and "clearness" which can be varied over time (Yamaha 2014: 150). This is reflective of the pop-genre emphasis of the Vocaloid singers, where the attitude of the sung text is arguably more important than a perfect rendering of the words. Vocaloid users debate whether the singer should attempt to sound human at all, with some users arguing that "if they sound like humans, they'll lose their distinctive quality" while other users contend that "human-like singing vocaloid will show the difference between good professional producers and just amateur producers" or that "a realistic vocaloid song show [sic] the capabilities of the software, and

helps to promote it" (Vocaloid Wikia 2014: web source). It is common to hear tracks produced within the Vocaloid fan community where sung text sounds stilted, with clear glitches between pitches, unintelligible lyrics, or incorrect dynamic emphasis on particular syllables (see Vocaloidotaku 2016). For these producers the primary goal is not necessarily to create an expressive rendering of human-like sung text, but rather to produce music which engages with the fan cultures surrounding Vocaloid singers and the Vocaloid platform. However, many of the Vocaloid singers are designed to have an effected, metallic timbre to their vocals, emulating the Auto-Tune effect quality found in many contemporary pop songs. This characteristic is evident in commercial Miku releases (see: Hatsune Miku 2016) which deliberately sound more robot than human. Yamaha have even embedded Vocaloid software into human-like robots such as the HRP4C for public performances (see Ikinamo 2009).

Virtual orchestral timbres and virtual vocal timbres have increased in realism and sophistication as the technology has developed. However, VSIs remain more vulnerable to claims of inauthenticity because the reproduction of the voice remains comparatively imperfect in the rendering of sung text. Users sometimes perceive a gap between the "composite of sound, visual representation, textual description and genre reference" of the VSI and the "actual experience" of using the product (Eidsheim 2009: web source). Musicians negotiate this gap either by selecting a different VSI to use, or by adapting their application of the VSI to accommodate them, as Morgan's VI research indicates (2015). The wide-spread use of VSIs and the growing audiences which listen to them suggest that the cultures of production and consumption surrounding VSI deployment are able to circumvent minor flaws in the design and application of the technology. The global reception of Hatsune Miku attests to this claim.

Eidsheim argues that there is a performativity of timbre encompassing performance as the "material creation of timbre" and "reception as performance" (2009: web source). For Eidsheim, performativity of timbre is generated through a kind of feedback loop between the performer and audience, whose preconceptions inform the singer and the singer responds by consciously or unconsciously shaping their vocal production (2009).[13] While Eidsheim is interested primarily in how this influences the connections between vocal timbre and race in the Vocaloid platform, the idea that musical timbre exists in a performative feedback loop between the sound producer/s and audience is useful to understanding the broader implications of VSIs as a site and practice of musical expression. In the case of Hatsune Miku, her simulated body and voice are a signification of an idealized Japanese teenage girl. Her constructed persona communicates particular cultural and identity significances for her audience. When users program Miku to sing, they work directly with her symbolic potential, communicating something between their own location, the lyrics, the composed and sounded music, and the listening economies in which the outputted music is circulated. Meaning is made from Miku's voice by the feedback loops and frictions created between these elements.

To understand how VSIs like Vocaloid might be conceived of as performative, we must first unpack the notion of performativity in more detail. Butler conceives of performativity as the way in which the anticipation of an essence, such as gender, "produces that which it posits as outside itself" (2014: xiv-xv).

Importantly, she considers performativity as repeated acts or rituals, which achieves "its effects through its naturalization in the context of a body, understood, in part, as a culturally sustained temporal duration" (Butler 2014: xv). Butler argues that coherence is desired in the construction of gender and that such coherence can be found in the acts, gestures, and desires which "produce the effect of an internal core or substance, but produce this on the surface of the body, through the play of signifying absences that suggest, but never reveal, the organizing principle of identity as a cause" (Butler 2014: 172-173). Butler further contends that "such acts, gestures, enactments, generally construed, are *performative* in the sense that the essence or identity that they otherwise purport to express are *fabrications* manufactured and sustained through corporeal signs and other discursive means" (Butler 2014: 173). To establish the validity of this claim, Butler analyses drag as an instance of gender parody, which reveals "that the original identity after which gender fashions itself is an imitation without an origin … a production which … in its effect – postures as an imitation" (Butler 2014: 175-176). In doing so Butler considers this perpetual displacement as constituting a "fluidity of identities that suggests an openness to resignification and recontextualization" (Butler 2014: 176).

Vocaloid voices can be considered a kind of drag act, playing out, through their sounding, the acts and gestures of human musical expressivity. Vocaloid voices placed inside robot bodies, or projected as 3D anime characters in concert tours, play at being human, dancing and singing inside their fabricated bodies for their audience with the gestures emblematic of J-Pop. The audience is privy to the fabrication, generating pleasure from the fact that Miku is not really a 16-year-old, 158cm tall Japanese girl. Like drag, all VSIs have the ability to antagonize some listeners because their sounding, which repeatedly produces acoustic devices associated with human musical performance, calls into question human musical performance by displacing it. The traces of VSI production – a metallic timbre, an imperfect legato, a glitch in sung text – reveal the imitation, but in doing so they also highlight the ongoing construction behind the production of such devices in human players, where the glitch is a variation that signifies humanity. Consequently, like gender, the notions of "humanness" and "expressivity" in musical performance can be understood as being continually resignified through the repetition of particular acts and gestures which are open to recontextualization.

Within Vocaloid fan communities there is debate around whether "realistic" "human-like" vocal qualities are even desirable. Some Vocaloid voices are designed with an audibly effected vocal timbre meaning that they cannot produce a "natural" sounding vocal quality. This is a different understanding to the ways MIDI orchestrators work with VSIs intended for classical music production, where a successful rendering of music is one where the listener never becomes aware that the singing was digitally produced. Instruments like Vienna Solo Voices and EWSC try to limit the possibility of producing inauthentic sounding renderings in the ways that they have constructed their sound libraries and through the interface design. The goal with these VSIs is to simulate humanness, rather than imitate it. In contrast, the Vocaloid interface design has less precision in its text-rendering than EWSC, but ample means of effecting the performance style of the vocalisation through the manipulation of Note Property control parameters.

Qualities like growling or breathiness, masculine or feminine vocal timbres, are part of the process of enacting human vocal expression via the repetition of its gestures. These gestures do not have to sound perfectly realistic to resignify the human vocal qualities they mimic.

Conclusion: music technologies, simulation and performativity

This article has unpacked the notion of human musical expressivity and explored how composers simulate these characteristics with the use of VIs. Further, it demonstrates that these instruments can be performative, playing out and destabilizing conceptions of human performance in collaboration with human users and audiences. Woody (2000: 14) says that musical expressivity possesses several sequential components incorporating inspiration of the composer, the produced musical text, the interpretive and expressive intentions of the performers, the sounding music, and the perception and emotional response of the listener. Similar processes are enacted when orchestral and vocal musics are realized with VIs. Software developers construct VI features to emulate the micro and macro sonic devices which signify human musical expressivity. In doing so they also sample human performers, amassing large libraries of individual performance articulations which can be deployed in varying and complex combinations. Composers and musicians employ VIs in creative ways which maximize the expressivity of such instruments. This is now embedded in the professional practice of MIDI orchestration and its associated pedagogies and user-communities which train composers in the realisation of expressive and human-sounding orchestral music. Related media industries such as film, television, and games reinforce composers' use of VIs by encouraging their adoption within their work cultures to an extent that the use of VIs has become normalized. Media employing VIs are transmitted to audiences who engage with them, and whose reactions or non-reactions feed back into VI development, use, and enculturation. The construction of VIs as devices of musical expressivity is, therefore, an evolving and mutually constructed endeavour.

VOIs are designed to simulate musical expression. When they produce enough of the characteristics of human orchestral or vocal music they call into question the distinction between the human and the virtual for the listener. However, it currently takes a skilled human composer-user to realize such effects through the crafting of sequences of musical gestures which render the music expressive.[14] In this sense, features like VSL's HPC anticipate the essence of expressive musical performance. The acts of programming, mixing, and otherwise refining a composition rendered by VIs are performative, in that they occur within user communities, fan communities and work cultures where such fabrications are manufactured and sustained through the reproduction of corporeal signs, such as the effected human musical expressions, and other discursive means, such as orchestration manuals and Facebook community discussions. In doing so, 'humanness' and 'expressivity' are being continually resignified through the repetition of particular acts and gestures which are open to recontextualization.

VIs like Vocaloid are refined in development cycles, which reinforce and refresh sonic gestures recurrently. This is akin to the process of perpetual displacement and resignification that Butler describes. New Vocaloid singers are released regularly and older singers like Miku are updated if they remain popular. Fan cultures surrounding Miku are a logical extension of her performativity which, when combined with the systemic mediatisation of J-pop culture, allows a virtual singer to displace the human pop star because "we are in the digital age!" (Crypton 2016: web source) and such displacement has been naturalized. Fan feedback contributes to Miku's development cycle, and her voice and persona evolve with each iteration. Consequently, Miku's existence as a "collaboratively constructed cyber celebrity" (ibid) points to a broader naturalisation where humans employ the virtual to produce the effect of human expression. While Miku may appear to be an outlier in terms of mainstream impact on the J-Pop community, the quieter dominance of Vienna Symphonic Library, or Hollywood Orchestra in the practice of screen composition, indicates that such naturalisations of VIs are occurring in other genres, albeit in different ways – simulating the human rather than imitating it via the adoption of human-like personas.

By recontextualizing the gestures of human musical expression, VIs like Vocaloid may also be subtly decentring human expressivity from musical experience. By playing out the acts and gestures of human musical expressivity Vocaloid singers, fashioning themselves on the human, risk revealing that notions of humanity may also be an ongoing production negotiated through the performance of expressive gestures. That such gestures, and their subtle ruptures, can become a site of pleasure for Miku's audiences points to a willingness to use VIs and related media forms as a space for experimentation with humanness rather than merely expression.

Endnotes

[1] Terminology referring to VIs is not consistently applied across the music industry (Morgan 2016: 4). VIs are also referred to as "software synthesizers", "softsynths", "software samplers", "plug-ins", "sound libraries", "software instruments"... This article has chosen to use the term "virtual instrument software" because it encompasses software which synthesizes, samples, and processes sounds, and is therefore more representative of increasingly complex and hybridized methods of realizing musical sound within software.

[2] There is significant scholarship interrogating the notion of expressivity in music performance, and it is beyond the scope of this article to provide an effective overview of such terrain. What is most relevant to this article is understanding the relationship between specific devices of human musical expression so that we can understand how these play out in the design and application of contemporary VIs.

[3] See respectively Garritan (2016), IK Multimedia (2016), UVI (2016), and Spitfire Audio (2016).

[4] Copyright Vienna Symphonic Library. Published with permission.

[5] The number of available patches wasn't clear from software documentation. This number was arrived at by counting the patches available within the software.

⁶ Copyright East West. Published with permission.

⁷ There is competing terminology on the topic, and "MIDI orchestration" can be referred to as "programming", "producing"... I have chosen to employ the term "MIDI orchestration" because: there are credible and widely referenced texts on the topic which employ this term; terms like "programming" are nonspecific, also referencing other kinds of programming such as writing computer code; and because the term "MIDI orchestration" is specific in its reference to MIDI-based orchestral instrumentation, a topic of this paper.

⁸ Keyswitching is a technique where MIDI notes outside the playing range of the instrument are used to trigger changes to performance articulations. For example, if an instrument can play a pitch range of C3-G6, the notes C0-C2 could be used to trigger changes to the articulation. See FIGURE 1.

⁹ Music composed for a specific sequence in a motion picture.

¹⁰ The number of available patches wasn't clear from software documentation. This number was arrived at by counting the patches available within the software.

¹¹ Copyright East West. Published with permission.

¹² This includes some reissues of popular voices who were redeveloped for release on new versions of Vocaloid and mobile platforms. The most complete lists of current Vocaloid voices are collated on Wikipedia (2016) and a Vocaloid wiki, Fanloid (2016).

¹³ Virtual voices raise issues of individual identity, gender, and race by virtue of the personas attached to the products and the underlying timbral quality of the samples the product reproduces--samples which leave audible traces of the human singers from which they were recorded. Eidsheim (2009) and Bell (2016) both note the specific ways race and gender are encoded into Vocaloid libraries.

¹⁴ Conceivably, such expressive renderings could be automated as VI technology progresses. HPC points towards this potentiality.

References

Bibliography

Ah-Software. 2016. Hiyama Kiyoteru. http://www.ah-soft.com/vocaloid/kiyoteru/natural.html; Accessed: 10 April 2016.

Arturia. 2016. V Collection 4. https://www.arturia.com/products/analog-classics/v-collection/overview; Accessed: 30 March 2016.

Baudrillard, J. 1994. *Simulacra and Simulation*. Ann Arbour, MI: The University of Michigan Press.

Bell, S.A. 2016. The dB in the .db: Vocaloid Software as Posthuman Instrument. *Popular Music and Society* 39(2): 222-240.
DOI: 10.1080/03007766.2015.1049041.

Bennett, S. 2009. *Computer Orchestration: Tips and Tricks*. Norfolk: PC Publishing.

Brown, A. 2015. *Music Technology and Education: Amplifying Musicality*. London: Routledge.

Butler, J. 2014. *Gender Trouble: Feminism and the Subversion of Identity*. London: Routledge.

Chanan, M. 1995. *Repeated Takes: A Short History of Recording and Its Effects on Music*. London: Verso.

Clarke, E. –

1995. Expression in Performance: Generativity, Perception and Semiosis. In J. Rink Ed. *The Practice of Performance Studies in Musical Interpretation.* Cambridge: Cambridge University Press.

1998. The Semiotics of Expression in Musical Performance. *Contemporary Music Review* 17(2): 87-102.

Crypton Future Media. 2016. Who is Hatsune Miku? http://www.crypton.co.jp/miku_eng; Accessed: 7 April 2016.

Davies, S. 2011. *Musical Understandings and Other Essays on the Philosophy of Music.* Oxford: Oxford University Press.

East West. –

2013. *Symphonic Choirs User Manual*. Hollywood, CA: East West Sounds Inc.

2015. *Hollywood Solo Instruments User Manual*. Hollywood: East West Sounds Inc.

2016a. East West/ Quantum Leap Hollywood Solo Instruments. http://www.soundsonline.com/hollywood-solo-series; Accessed: 28 March 2016.

2016b. East West/ Quantum Leap Hollywood Orchestra. http://www.soundsonline.com/Hollywood-Orchestra; Accessed: 28 March 2016.

2016c. East West/ Quantum Leap Symphonic Orchestra. http://www.soundsonline.com/Symphonic-Orchestra; Accessed: 28 March 2016.

2016d. Home. http://www.soundsonline.com/; Accessed: 28 March 2016.

Eidsheim, N. 2009. Synthesizing Race: Towards an Analysis of the Performativity of Vocal Timbre. *TRANS: Transcultural Music Review* 13. http://www.sibetrans.com/trans/articulo/57/synthesizing-race-towards-an-analysis-of-the-performativity-of-vocal-timbre; Accessed: 7 April 2016.

Evans, A. 2005. *Sound Ideas: Music, Machines, and Experience.* Minneapolis, MN: University of Minnesota Press.

Fanloid. 2016. List: Official Vocaloid Characters. https://en.wikipedia.org/wiki/List_of_Vocaloid_products; Accessed: 10 April 2016.

Frith, S. 1996. *Performing Rites: Evaluating Popular Music*. Oxford: Oxford University Press.

Garritan. 2016. Garritan Instant Orchestra Sound Library. http://www.garritan.com/products/instant-orchestra/; Accessed: 30 March 2016.

Gilreath, P. 2004. *The Guide to MIDI Orchestration*. Marietta, GA: MusicWorks Atlanta. Third edition.

Godlovitch, S. 1998. *Musical Performance: A Philosophical Study*. London: Routledge.

IK Multimedia. 2016. Miroslav Philharmonik 2. http://www.ikmultimedia.com/products/philharmonik2/; Accessed: 20 March 2016.

Izhaki, R. 2009. *Mixing Audio: Concepts, Practices, Tools.* Oxford: Focal Press.

iZotope. 2016. Iris 2. https://www.izotope.com/iris; Accessed: 30 March 2016.

Katz, M. 2004. *Capturing Sound: How Technology Has Changed Music*. Berkeley, CA: University of California Press.

Kendall, R.A., and Carterette, E.C. 1990. The Communication of Musical Expression. *Music Perception* 8(2): 129-164.

Klein, E. –
  2014. Is Classical Music "Boring"? A Discussion of Fidelity, Virtuosity and Performance in Classical Music Recording. In J. O'Regan and T. Wren Eds. *Communities, Places, Ecologies: Proceedings of the 2013 IASPM-ANZ Conference*, Brisbane, Australia, 24-26 November 2013: 112-125.
  2015a. Opera Composition and Performance Utilising Computer-based Recording Technologies and Virtual Instruments. In T. Gifford Eds. *Proceedings of Createworld 2015*. CreateWorld 2015, A Digital Arts Conference, Brisbane, Australia, 12-13 February 2015: 10-17.
  2015b. Performing Nostalgia On Record: How Virtual Orchestras and YouTube Ensembles Have Problematised Classical Music. *Journal on the Art of Record Production* 9. http://goo.gl/EOQm3O; Accessed: 27 August 2016.

Kopiez, R., Wolf, A., Platz, F., and Mons, J. 2016. Replacing the Orchestra? The Discernibility of Sample Library and Live Orchestra Sounds. *PLoS ONE* 11(7): e0158324. DOI 10.1371/journal.pone.0158324; Accessed: 29 December 2016

Juslin, P.N. 2000. Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception. *Journal of Experimental Psychology: Human Perception and Performance* 26(6): 1797-1813.

Lindström, E., Juslin, P.N., Bresin, R. and Williamon, A. 2003. "Expressivity Comes from Within Your Soul": A Questionnaire Study of Music Students' Perspectives on Expressivity. *Research Studies in Music Education* 20(1): 23-47.

Local 802. 2003. Local 802 News – Publications & Press Releases. http://www.local802afm.org/publication_entry.cfm?xEntry=3954395; Accessed: September 2011.

Morgan, C. –
  2014. Screen-music Orchestration Quantified: Preliminary Results of a Study on the Use of Virtual Instruments. Music and the Moving Image Conference, New York University, Steinhardt, May 2014. https://www.academia.edu/7926743/MAMI_Presentation_-_Craig_Morgan_20140509; Accessed: 3 April 2016.
  2015. Virtual Instruments and the Future of Australian Film Music Composition. National Conference of the Musicological Society of Australia 2015. University of Sydney, July 2015. https://www.academia.edu/16493620/Virtual_Instruments_and_the_future_of_Australian_film_music_composition: Accessed: 5 April 2016.
  2016. *The Use of Virtual Instruments by Australian Screen Composers*. Sidney: University of Sydney. http://goo.gl/SS9Oai; Accessed: 27 August 2016. PhD Thesis.

Native Instruments. 2016. Rounds. https://www.native-instruments.com/en/products/komplete/synths/rounds/; Accessed: 25 March 2016.

Palmer, C. 1989. Mapping Musical Thought to Musical Performance. *Journal of Experimental Psychology: Human Perception and Performance* 15(12): 331-346.

Pätynen, J. and Lokki, T. 2016. Concert Halls with Strong and Lateral Sound Increase the Emotional Impact of Orchestra Music. *The Journal of the Acoustical Society of America* 139(3): 1214-1224. DOI: 10.1121/1.4944038.

Pejrolo, A. 2011. *Creative Sequencing Techniques for Music Production*. Oxford: Elsevier. Second edition.

Pejrolo, A. and DeRosa, R. 2007. *Acoustic and MIDI Orchestration for the Contemporary Composer.* London: Focal Press.

Player, T. 2014. Real Orchestra Vs. Synth Mockup – Part 1/6. http://audiocookbook.org/real-orchestra-vs-synth-mockup-part-16/#more-11013; Accessed: 27 August 2016.

Sh-henian. 2016. Luo. http://www.sh-henian.com/idols/lty; Accessed: 10 April 2016.

Spectrasonics. 2016. Omnisphere 2. https://www.spectrasonics.net/products/omnisphere/; Accessed: 10 April 2016.

Spitfire Audio. 2016. Albion One: Epic Composer Tools.

http://www.spitfireaudio.com/shop/ranges/albion/albion-one/; Accessed: 1 April 2016.

Sundberg, J., Filipa M.B. and Himonides, E. 2012. Intonation and Expressivity: A Single Case Study of Classical Western Singing. *Journal of Voice* 27(3): 391.e1-391.e8.

Symes, C. 2004. Setting the Record Straight: A Material History of Classical Recording. Middletown, CT: Wesleyan University Press.

Talk Classical. 2011. Sound Synthesizers – Future of Classical Music? http://www.talkclassical.com/13084-sound-synthesizers-future-classical.html; Accessed: 27 August 2011.

UVI. 2016. Orchestral Suite. http://www.uvi.net/en/orchestral-composer/orchestral-suite.html; Accessed: 14 March 2016.

Vienna Symphonic Library –

  2015. *Vienna Instruments Pro User Manual 2.5*. Vienna: Vienna Symphonic Library.

  2016a. Company. https://vsl.co.at/en/Company; Accessed: 26 March 2016.

  2016b. Products. https://vsl.co.at/en/Products; Accessed: 26 March 2016.

  2016c. Vienna Instruments Pro. https://vsl.co.at/en/Vienna_Software_Package/Vienna_Instruments_PRO#!Product_Info; Accessed: 26 March 2016.

  2016d. Vienna MIR Room Pack Bundle. https://vsl.co.at/en/Vienna_Software_Package/Vienna_MIR_RoomPack_Bundle; Accessed: 2 April 2016.

  2016e. Vienna Solo Voices. https://www.vsl.co.at/en/Voices_Complete/Vienna_Solo_Voices; Accessed: 12 April 2016.

  2016f. Vienna Super Package. https://vsl.co.at/en/Instruments/Super_Package; Accessed: 12 April 2016.

Vocaloidotaku. 2016. Songs & PVs & MMDs.
    http://www.vocaloidotaku.net/index.php?/forum/11-songs-pvs-mmds/;
    Accessed: 1 September 2016.
Vocaloid Wikia. 2014. Robotic or not Robotic, that's the Question...
 http://vocaloid.wikia.com/wiki/Thread:73705; Accessed: 1 September 2016.
Walden, J. –
    2004a. Vocaloid Leon & Lola. *Sound on Sound*. March 2004.
        http://www.soundonsound.com/sos/mar04/articles/vocaloidlandl.htm;
        Accessed: 7 April 2014.
    2004b. Vocaloid Miriam. *Sound on Sound*, December 2004.
        http://www.soundonsound.com/sos/dec04/articles/miriam.htm; Accessed: 7
        April 2014.
Wikipedia. 2016. List of Vocaloid Products.
    https://en.wikipedia.org/wiki/List_of_Vocaloid_products; Accessed: 10 April
    2016.
Woody, R.H.
    1999. The Relationship Between Explicit Planning and Expressive Performance
        of Dynamic Variations in an Aural Modeling Task. *Journal of Research in
        Music Education* 47(4): 331-342.
    2000. Learning Expressivity in Music Performance: An Exploratory Study.
        Research Studies in Music Education 14(1): 14-23.
Yamaha –
    2014. Vocaloid 4 Editor User Manual. USA: Yamaha.
    2016. Vocaloid. http://www.vocaloid.com/us/; Accessed: 6 April 2016.

Videography

Hatsune Miku. 2016. Blue Star feat. Hatsune Miku.
    https://www.youtube.com/watch?v=H_6I_LLheAY; Accessed: 1 September
    2016.
Ikinamo. 2009. Yamaha Vocaloid Robot Singing Using Speech Synthesis
    Software. https://www.youtube.com/watch?v=mfxkhzGqZIs; Accessed: 1
    September 2016.