Affective Sound Synthesis: Considerations in Designing Emotionally-Engaging Timbres for Computer Music

Aura Pon[a], Dr. David Eagle[b], and Dr. Ehud Sharlin[c]

(a) Interactions Laboratory, University of Calgary, aapon@ucalgary.ca
(b) Music Department, University of Calgary, eagle@ucalgary.ca
(c) Interactions Laboratory, University of Calgary, ehud@cpsc.ucalgary.ca

Abstract

In this paper, we examine the essential role of instrument timbre in communicating and eliciting emotion in music, and consider how this knowledge can inform the design process of sound synthesis towards achieving maximum expression. In the composition and performance of acoustic music, timbres are carefully chosen to capture and communicate the emotional undercurrent of the musical material. Instrumentalists master a diverse palette of tone-colours to increase their versatility in interpreting meaning and emotion in their music, and composers painstakingly consider instrument choice in search of the best voice to express their ideas. We believe that sound in computer music, with its infinite timbral choices available through digital sound synthesis, requires at least equal (if not more) attention and consideration towards emotional quality and depth. As we undergo a process of designing original sounds that will be included in *Vuzik*, our child-friendly digital music composition interface, we explore what aspects of sound contribute to its emotional character. We examine how humans perceive emotional quality in sound and its relation to our use of tone in speech to convey emotion, and how to consider emotional engagement in our selection of acoustic attributes to emphasize in the creation of original digital timbres. We relate to these topics through our experiences in attempting to design timbres for the *Vuzik* interface that will be engaging to a variety of users. We also examine reasons behind complaints of "emotional coldness" in computer generated sounds, and relate them back to lessons we can infer about the emotion conveyed by their timbres. Through these inquiries we aim to illuminate new considerations to be incorporated in sound synthesis, and hope to elevate the capabilities of musical expression with digital timbres.

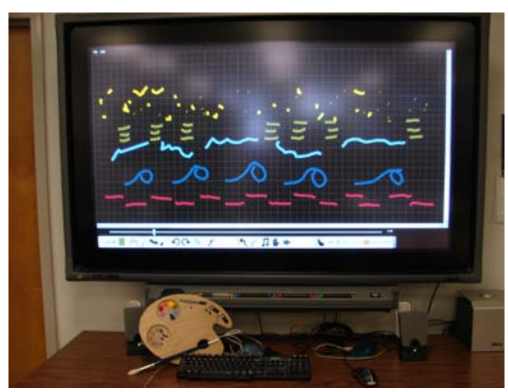
Introduction

Musical composer Rimsky-Korsakov once asserted that the tone colours of a musical work are the very soul of the music, an essence inseparable from the music's content [5]. Human fascination with the wide array of diverse sounds around us, and our keen ability to identify these sounds, have naturally led to sound timbre being an essential attribute of aural communication in both speech and music. Timbre is essential in human use of vocal tone quality to express emotions. To similarly harness the power of different qualities of sound to communicate emotion in music, the art and science of orchestration were developed, where composers learn how to paint with the many subtle instrument colours, or timbres, to capture and communicate the emotional undercurrent of their musical material. Until relatively recently, composers drew their timbres from acoustic musical instruments whose forms often evolved over centuries. Instrument builders throughout history applied complex wisdom about how to physically construct a generator of sound that delights human ears. With the advent of computer-generated sound, tones were freed from constraints of the physical forms of acoustic instruments and thus almost infinite timbral possibilities emerged. Certain questions arise from this. How does one wishing to design an original sound choose from such a huge array of sound attributes? What do these modern day instrument designers have to consider, and what can be learned from acoustic instrument builders? Ultimately, the goal of the digital instrument designer is the same as that of the acoustic instrument builder, as is the same as the performer and the composer - to evoke listener ecstasy. As we approach the role of digital instrument builder for a music composing interface called *Vuzik* [4], we examine what we think needs to understood about timbre and the role it plays in music, what we can learn from in the sounds of acoustic instruments, and what possible pitfalls to avoid in digital sound synthesis, towards a goal of creating engaging new timbres. In this position paper, we explore and brainstorm some possible considerations for digital sound synthesis, pursuing sound which will be more emotionally charged and engaging for use in the design of future computer musical instruments.

Background Motivation

Our motivation for discovering strategies for synthesizing engaging digital timbres arises from the design efforts of *Vuzik* (Figure 1), an interface for composing music by painting on an interactive surface *[4]*. Although ideally appealing to a variety of users, we wanted above all for this instrument to be engaging to children, and inspire in them some level of understanding and interest in creating their own music. *Vuzik* functions as a compositional canvas, where the visuals, which are created by tangible tools such as a paintbrush and palette, are mapped to corresponding attributes of sound. The *Vuzik* visuals are in turn serving as a performance instrument when the composition is played back.

Figure 1 – Vuzik Interface



Thus, the sounds available to the creator are an integral part of this interface. The General MIDI timbres used for an early prototype of *Vuzik* were found to be unsatisfactorily engaging for reasons we would like to identify, and it seemed unidiomatic to use instrument sounds derived from acoustic instruments given the differing form and digital nature of *Vuzik*. Thus we thought it important to synthesize original digital timbres for our interface. In an early design critique of Vuzik, performed with a small group of children, one of our participants, an 8 year old user requested for the inclusion of sounds she recognized, like the Ukulele she played. We realized that for the interface to be engaging for musical composition and performance, there were many considerations for its available sounds, such as whether they are relatable, complex, contrasting, and alluring. Ideally, we would like the sounds of *Vuzik* to at least match an acoustic instrument's ability to engage a young listener with its sound quality. We will attempt to convey here what knowledge of audition, acoustic instruments, and synthesis techniques are informing us in the process of creating eight engaging sounds for this new digital instrument.

Timbre Perception

In trying to create timbres that are engaging to people, it follows one must consider how humans perceive and understand timbre, and what attributes about timbre affect them mentally and in what way. We wish to highlight here certain aspects of human perception of timbre that are relevant to consider in designing timbres.

Every tone is made up of a fundamental frequency, and many higher frequencies, called overtones. The relative intensities of these overtones, and the pattern of their onset and decay over time, determine the unique, distinctive character of each sound, or the timbre. Levitin compares this unique pattern that composes timbre as the sound generator's *timbral fingerprint* [3]. Tones which are said to have a definite pitch typically have overtones with frequencies related by simple integer multiples of the fundamental frequency; the degree to which there is a lack of orderly relationship between overtones and fundamental frequency that the brain can detect is the degree to which a sound will be heard as simply noise. Also, overtone pattern determines the brightness or darkness of the sound, in terms of the strength of high frequency overtones versus low frequency overtones, respectively.

The ability to identify different timbres is perhaps the most important and ecologically relevant to survival of our auditory capabilities. Not only has it been essential for distinguishing sounds of danger from sounds associated with the basic necessities of life such as a food source, but with the importance of language and vocalization in the complex social nature of human beings, vocal timbre has been one important vehicle for individual and affective expression. Thus, timbral discrimination in humans is so highly developed that many people can recognize hundreds of different voices, and discriminate between many subtleties of emotion in a familiar person's voice. [3] Numerous studies show people can identify emotional states from the voice alone with an accuracy of 50-60%, as compared to 12% accuracy from random guessing, and that timbre is proven to be one quality of the voice that plays a definitive role in affective expression [7].

With our timbral recognition being highly evolved towards the role it plays in language, this affects the types of sounds to which we are most engaged by. One notable example of this is the tendency for our ears to be most sensitive to higher tones in the same ranges as speech [2].
One attribute of a tone that is essential to our ability to characterize and distinguish timbres is the *envelope* of the sound, i.e., the rate, intensity, and overtone pattern of the attack, decay, sustain, and release of a sound. Though the attack/decay phases, where energy is first applied to an acoustic instrument, contains a large element of noise, it is likely one of the most crucial parts of the character of the sound, as shown by experiments conducted by Pierre Schaeffer in the 1950s [6] where most people have shown difficulty identifying certain instruments like a trumpet or flute when the attack is removed.

The envelope of a sound is most related to how the sound changes over time, which points at what we feel is the most salient auditory consideration: the brain being ultimately sensitive to change. As Jourdain reports approximately 85% of primary auditory neurons exhibit a phenomenon called habituation, where, the longer these neurons are stimulated in the same way, the less they respond. Without constant renewal of a sound (or renewal of attention to sound), we deafen to it. Thus, if the sound has more minute variations within it, of timing and loudness and pitch, etc., it would be more intense and engaging. Jourdain provides an example of this comparing an acoustic instrument to a synthesized one that is particularly germane to this paper: "A real violin tone is 'richer' than a synthesized one partly because it gives auditory cortex more to respond to; a violin vibrato is richer still." [2]

From the aforementioned reflections, we conclude that this *richness* in sound that engages people can be created through promoting complexity and relationships in the overtone fingerprint of a sound, learning from the affective use of vocal tone in language to give emotional quality to a timbre, and the incorporation of complexity of change and variation in the sound.

Lessons from Acoustic Instruments

Acoustic musical instruments have been, and still remain, the primary source of sound and colour in music. The craft of instrument building has been honed over centuries to bring forth a variety of sound generators that create sounds that are engaging to listeners. Instruments have evolved or gone extinct in a type of natural selection where the criterion of survival is what communicates effectively to human ears in various ecological contexts. From the starting point of these instruments, musicians further perfect their use of them towards maximum expression in music. Instrumentalists master a diverse palette of tone-colours to increase their versatility in interpreting meaning and emotion in their music, and composers painstakingly consider instrument choice, combination, and register in search of the best voice to express their ideas. We identify some strengths of acoustic instrument sound that can be applied to digital sounds to make them more engaging.

The effectiveness of acoustic instrument timbre originates from its complexity and variation. Our ears particularly respond to complex sounds, which is why humans have designed acoustic musical instruments that resonate across a wide range of frequencies in bands called formants. Because an acoustic instrument's formants remain the same for high notes and low notes alike, sometimes the formants emphasize a tone's fundamental frequency, while on other occasions the formants may emphasize the low harmonics and high harmonics [2]. This is why an acoustic instrument sounds different across its range; the overtone patterns vary with pitch and loudness over the different registers. For example, the soft velvety tone of a flute in its lowest register is very different from the pure, bright, piercing sound in its upper register. High notes tend to be poor in overtones because instruments have trouble vibrating at very high frequencies. Bass tones tend to be muddy in many instruments since their higher overtones tend to fall in the frequency range at which our ears are most sensitive. These different registers are still identified as originating from the same single instrument, but there is noticeable difference in the timbre. Composers capitalize on this registral difference by scoring an instrument in the register with the timbre that they believe is most expressive for their musical idea and expression. This subtle variation in timbre also creates interest and stimulation for listeners' ears as the instrument moves through various notes of its range.

Another parameter in acoustic instrument sound that displays complexity and variation is the nature of the attack of the sounds. The attack of sounds varies greatly between different instruments, which creates very contrasting, different timbres to use in orchestration, and large degrees of emotional expression. The articulation of a note with a trombone differs greatly from that a cello. As well, each instrument can achieve a variety of articulations that will vary in quality with their registers.

It is worth noting the expressive control of the envelope of each tone produced by some acoustic instruments; again, another instance of complexity and variation. Most sustaining instruments can subtly vary of the pitch (such as in vibrato), the intensity, amplitude, and tone quality of each note for expressive effect. An oboe or a violin can achieve great variance in the "shape" of a single tone over time.

Lastly, human imperfections or irregularities in how musicians play each tone, whether intentional or unintentional, is always a part of a live performance of any acoustic instrument. This in itself contributes to much complexity and variation in acoustic sound, which adds interest and stimulation for the brain.

All these subtle variations in acoustic sound culminate into a sound that is *alive* – dynamic, flexible and expressive. We feel this dynamism is essential for expressive verbal communication and music alike.

Early Reflections on Digital Sound

In our implementation of the *Vuzik* prototype, we needed to select eight instruments timbres to be represented by the colours on the palette. We initially selected these timbres from the General MIDI Program bank, aiming for contrast among the sounds and attempting to avoid sounds that imitated acoustic instruments. Although the sounds served the basic purpose of allowing young users to interact with *Vuzik* using eight unique sounds in the preliminary design critique, we felt the timbres lacked strong character and interest, and that it was difficult to find timbres that had differing characters from one another. Although this initial study did not allow us to isolate the effect the timbres had on user engagement levels, we hypothesize that if the sounds were indeed indistinct, uninteresting, and too simple, as they seemed to our ears, then the aural engagement with these sounds would not match the visual engagement that users had with of the bright, contrasting colours representing the sound. This would affect the success of the interface in inspiring users to explore sound and music composition. As we began to explore frequency modulation (FM) synthesis as an alternative source of new timbres, we sought to understand what was not desirable about the General MIDI timbres so as to avoid its pitfalls.

In general, the MIDI timbres seemed cold, static, artificial, and bland. One possible concern that could lead to this problem is whether the sounds contain a complex timbral fingerprint in terms of the number and strength of overtones they display. In reflecting on observations put forth about audition and acoustic instruments previously, we suggest the main problem lies in a lack of variation and change. Through our familiarity with acoustic instrument sounds, as well as that of our own voices, we come to expect some of these variations we described above, such as subtle changes in tone quality with register, radically different envelopes and attacks with each instrument as well as in different registers. Perhaps when we hear a digital instrument that lacks these variations, we feel it is “unnatural” or “artificial” and thus cannot relate to it as an expressive musical instrument to the same extent as to an acoustic instrument. Such regularity and uniformity of certain parameters from one pitch to the next, and from one instrument to the next, is something rarely encountered in human voice or in acoustic instruments.

We recognize the need to overcome these challenges when working with FM synthesis if we are to attempt to create engaging timbres for *Vuzik*.

Considerations for Designing Digital Sounds

We are hoping to gain insight from our examination of audition and the strengths of acoustic sound above to develop an awareness of the considerations needed to synthesize new sounds that are dynamic and interesting. However, we also want to take full advantage of the new affordances of digital sound synthesis beyond any limitations of acoustic instruments and achieve unique timbral qualities that are unlike existing acoustic instruments. We suggest some considerations we are incorporating in our digital synthesis processes based on our aforementioned reflections.

- **Complex timbral fingerprints**, where each sound contains a unique and rich pattern of overtone intensities, will contribute to the distinctiveness and depth of the sound. Awareness of the effect of the strength of high frequency overtones versus low ones on the brightness or darkness of sound should be taken into account.
- **Consideration to vocal ranges and the role of tone quality in vocally conveying emotion** could help inform the design of sounds for particular emotional qualities and ensure greater listener sensitivity.
- **Distinctive shaping of the sound envelope**, with particular attention to the attack of the tone and its incorporation of noise to give it a physical quality, will contribute to a strong, contrasting character in the timbre.
- **Variation in various parameters of the sound** may help keep the brain engaged by continuously renewing its attention to the sound through subtle change. Subtle change and

variation can be incorporated into parameters such as the overtone patterns, envelopes and attacks for different registers of the instrument. Within each tone, the rate of frequency modulation (*vibrato*) or the contour of envelopes of the sounds could vary depending on the duration or amplitude of the tone, or other factors.

- **Human gestural control** of some of these aspects of timbre and variation, if logical for the interface, can create a more dynamic, responsive, and expressive sound that may help to keep the person interacting with it engaged.

Conclusion

The introduction of new digital musical interfaces such as *Vuzik* opens up new sonic and interactive possibilities and opportunities to engage people in unique ways not previously available. The success of *Vuzik* in inspiring people, especially young users, to explore music could be enhanced by ensuring that its sounds are designed to be engaging and dynamic. With the added capabilities of digital sound synthesis to create sounds for such an interface come an almost overwhelming number of choices of sound attributes and timbral character. We have examined how knowledge of how humans hear and discriminate timbre, its relationship to vocal quality in affective verbal communication, and particularly awareness of the brain's sensitivity to change, can help inform our design of *Vuzik* sounds to be more effectively alluring the human ear. We identified certain strengths displayed by the sounds of acoustic instruments that could also be considered by the digital instrument designer. With the increasing number of choices available in the creation of digital timbre, principles and guidelines for what will effectively excite the human ear and convey an array emotions in music would be exceedingly helpful. Through further efforts to discover and incorporate such principles into our digital sound synthesis process, we hope provide *Vuzik* and other digital instruments with the "very soul of music" that is brought to life through timbral colour.

References

[1] Collins, Nick. (2010). *Introduction to Computer Music.* UK: John Wiley & Sons Ltd.

[2] Jourdain, Robert. (2002). *Music, the Brain, and Ecstasy: How Music Captures our Imagination.* USA: Quill.

[3] Levitin, Daniel J. (2007). *This is Your Brain on Music.* UK: Plume.

[4] Pon, A., Ichino, J., E. Sharlin and D. Eagle. *Vuzik: Music Creation and Comprehension through Painting*, Technical report, Computer Science Department, University of Calgary, 2011-994-06, February 11, 2011, https://dspace.ucalgary.ca/handle/1880/48421

[5] Rimsky-Korsakov, Nicolay. (1933-1983). *Principles of Orchestration.* New York : Kalmus.

[6] Schaeffer, Pierre. 1968. *Traites des objets musicaux.* Paris: Le Seuil.

[7] Scherer, Klaus R. (1995). *Expression of Emotion in Voice and Music.* Journal of Voice. USA: Lippincott Raven Publishers, Vol. 9, No. 3, pp. 235-248.