Timing, Forget the Timing

James Lemin

School of Computing, Engineering and Physical Sciences
University of Central Lancashire

JLemin@uclan.ac.uk

BSc (Hons) Computing (Year 1)

ABSTRACT

Computers are increasingly being used for the composition and production of music. It is entirely possible for composers to compose, produce, master and release their music without ever leaving their computer. This paper questions whether trusting the performance of music to a "humanize" algorithm results in an aesthetically less appealing result. To do this, 35 subjects were asked to appraise a range of sound files produced with varying levels of a humanize preset. The clear result was that the application of the humanize algorithms had little to no effect on listener's experience of the music.

Keywords

MIDI, Timing, Humanize, Quantize, Sequencer, DAW, Music.

1. INTRODUCTION

Music composition using computers is big business. The National Association of Music Merchants (2014) reported that in 2013, the combined sales of notation, and recording and sequencing software accounted for over $160m dollars in the US alone. Notation and Digital Audio Workstation sequencing software (DAW) allows composers to not only write music, but also render a definitive "performance" of the piece of music. Due to this, the traditionally clearly defined line identified by Cone (1987) between composition, and performance has become increasingly blurred. Moreover, the additional disconnect identified by Lerdahl (1992: 101) between the "production and consumption" further removes the listener from the element of performance.

Cone (1987) identifies rhythm as the keystone to music, whilst a single sound can be musical, it is not, in itself music. As the humanize function in Logic explicitly alters the timing, and by extension the rhythm of events, one could reasonably expect research in this area to yield greatest dividends.

Apple (*2009*) describes the humanize function available in its' flagship DAW Logic Pro as something that "Adds a random value to the position, velocity, and length of selected note events". The default setting alters events by up to 10%. It is inferred that a user, with just a couple of clicks, can "add life to compositions. This seems too good to be true; can the additional "life" introduced by the humanize algorithm quantifiably improve the listener's experience of the music, or does the appreciation of a piece of music rely on other factors?

This paper depicts an experiment to explore the issue surrounding the humanization of music and its appeal to participants. The hypothesis for this experiment is that subjects will prefer subtly humanized files over the unaltered original file.

2. DESIGN

The experiment will be broadly split into two sections. The first part of the experiment is concerned with the piece of music and how the humanize algorithms changed it. The requirements for the chosen piece of music are quite specific; firstly, the music has to be recognisable so participants are likely to be familiar with it and could reasonably be able to judge it. To enable listeners to more easily hear the effects of the humanize algorithms, the piece should also be monophonic, in a simple time signature, and the piece should be played on an instrument with a short attach and long sustain in order to best represent the note length and start position.

3. PROCEDURE

3.1 Applying the Humanize Preset

With these parameters in mind, a section of Rimsky-Korsakov's Flight of the Bumblebee was chosen, and the main motif was transposed into midi as a monophonic representation of the score. This initial file was saved and used as a baseline for the experiment.

To survey the range of the humanization preset, this baseline file was be duplicated, Logic default humanize preset applied and saved as a midi file. The humanize function in Logic defaults to altering both timing and velocity of notes. In order to examine the effects of timing alone, the changes to velocity were disregarded.

To analyse the resulting midi files, they were de-compiled into comma-separated values. The start and end point of each note was recorded and compared with the original event. After this process had run fifty times, five files were chosen that represented the overall minimum (greatest negative), first quartile (Q1), median, third quartile (Q3), and maximum (greatest positive) deviations from the baseline file. These five files, alongside the baseline file were rendered to audio using the same settings to ensure parity

3.2 Measuring Responses

The second section of the experiment looked at the response to the subtle timing changes introduced by the humanize process. To do this, quota sampling was used to ensure the results could be applied to the larger student population in general. Whilst a rigorously implemented random sample would have been preferable, it was not possible to identify full population in order to produce a random sample. This being the case, Peterson and O'Dell's (1950; 182) pragmatic approach of "a less costly quota sample may meet the accuracy specifications" was adopted. Respondents were asked to identify their favourite and least favourite renderings of the piece of music as well as comment on how obvious they found the differences between the renderings.

4. RESULTS

4.1 Midi File Analysis

Apple's default Humanize preset states that the range of difference should be +/-10%. This was borne out by the analysis of the humanized midi files that proved that over 99.5% of events were within of 10% of the original values. Within this +/-10% range, there was an even spread; across this range, the median and mean values settled at almost exactly 100%. This again gives further credence that the algorithm does arbitrarily move events around in a random way with a range of +/-10% irrespective of meter, pitch or any other factor.

4.2 Responses to Midi Files

Using figures obtained from the Higher Education Funding Council England (2013), a quota of thirty-five students based on age, gender and level of study was chosen to represent the wider student population at the University of Central Lancashire (UCLan). Of this sample, 30 (85.7%) had heard the piece of music before. Of those who hadn't previously heard the piece before, all were under 21.

Irrespective of their choice, respondents who were able to choose a favourite and least favourite track reported that the differences were not very pronounced.

Table 1. Responses to MIDI files

	Favourite	Least Favourite
None	30	30
Baseline	2	0
Min	3	0
Q1	0	0
Median	0	3
Q3	0	0
Maximum	0	2

5. ANALYSIS AND DISCUSSION

There are many possible reasons as to why the respondents did recognise the differences between the rendered pieces of music. The first is that the differences between the pieces were just too subtle to be obvious to the majority of listeners.

Human ability to perceive events at the micro scale may also have played a part. Dennis Gabor (1947) was the first to prove that listeners typically require between 10ms and 21ms to discern a note. Across the five selected files, the maximum difference observed was 125ms. This 125ms delay affected the start of the first note, so was essentially imperceptible. Across the remaining notes, the values were all within the range of +/-20ms, the majority of which were within +/-10ms. When combined with the attack of the piano chosen (around 5ms) it is understandable that respondents were simply unable to perceive the difference between the different renderings. Something else may have been at play however. The human ability to hold a tune has been well documented Kubitz (2010) recounts a story of Galileo who, faced with the lack of accurate timing instruments, hummed a tune to himself as means of accurately measuring time for his experiments. Sacks (2013) argues that this process of entrainment is responsible for listeners not only "keeping time" but anticipating patterns. If this is the case, then minute variations to the timing of events may be "corrected" by the listener. Sessions (in Cone, 1987: 139) reported a similar phenomenon; listeners remake music in their own imagination

Perron (1994). reported that deviations from the required tempo by a mean average of 3.5% were largely missed by listeners, even by professional musicians. The following year, Letivin and Cook (1996) found that whilst long term memory of tempo was remarkably precise, there was still a margin of error of 4% for the majority of test subjects. The deviations from the tempo recorded in this experiment were typically significantly smaller which goes some way to explaining why the majority of subjects could not hear a difference between the files.

An additional possible reason for subjects being unable to discern differences in the sound files may be related to the abstraction of the experiment itself. Subjects were asked to appraise pieces of music that were devoid of all spectacle of performance. Experiencing music on headphones is vastly different to the spectacle of a live performance of the same piece of music.

6. CONCLUSIONS

Whilst the concept of making a composition sound less mechanical by subtly altering the timing has merit, it is clear that the effect was lost on the majority of respondents. To understand this better, Jackendoff and Lerdahl's work in 2006 linking the grammar of music and the grammar of language is a useful metaphor; subtly altering the timing of speech patterns does not alter the perception of the words or overall message. By the same token however, it can't reasonably be expected to add any element of "performance" to the piece of music.

What differentiates a computer rendering of a composition from a live performance must surely be the interpretation and performance of the musician/s. The idea of a two-click computer preset that is able to humanize a composition performance is deeply flawed, particularly when all the preset does is to move events around by up to 10%. Musicians tend not to be inaccurate in a random way; Tempereley (2007: 26) reports that performance "errors" tend to reinforce the musical meter: "metrically strong notes tend to be played slightly more legato and louder than others", and when performers play a note in the wrong place, the played location tends to be a beat of similar metrical strength to the correct one"

What defines a good performance is a subjective cocktail of different factors; the performer, venue, instrument, audience and so on all combine to make each performance unique. Given this level of complexity, and the individual nature of appreciation of art, it is highly unlikely that a computer could successfully model enough of these factors to pass a musical equivalent of the Turing test, but there is clearly scope for improvement over a humanize preset that arbitrarily pushes events around based on a random number.

7. FUTURE WORK

The focus of this study was, by design, narrow. Further investigation into the effects of timing on other pieces of music, particularly polyphonic music may prove fruitful.

Whilst this study concerned itself with the timing effects of the humanize preset, the effects of velocity (how loud or soft a note is played) were disregarded. Judging by the results presented here, Velocity is likely a better measure of expressive performance than timing, so would be an obvious next step. Further research into what factors contribute to the aesthetic appreciation of a performance would also be useful in identifying themes that software could model.

8. REFERENCES

Cone, E (1987) “Music and Form” in Alperson, P. (ed.) *What is Music?.* Pennsylvania: Pennsylvania state university press. pp.133-146.

Gabor, D. (1947) Acoustical Quanta and the Theory of Hearing. *Nature*, 159(4044), pp.591-594.

Higher Education Funding Council of England (2013). *HEFCE Equality Data Overview : Students studying at English HEIs*. [online] pp. 1. Available at: http://www.hefce.ac.uk/media/hefce/content/whatwedo/leadershipgovernanceandmanagement/equalityanddiversity/equalities_summary_student.pdf [Accessed 6 Feb. 2015].

Jackendoff, R. and Lerdahl, F. (2006) The capacity for music: What is it, and what’s special about it?. *Cognition*, 100(1), pp.33-72.

Kubitz, A. (2010) *The Elusive Notion of Motion: The Genius of Kepler, Galileo, Newton, and Einstein*. Dog Ear Publishing, pp.69.

Lerdahl, F. (1992) Cognitive constraints on compositional systems. *Contemporary Music Review*, 6(2), pp.97-121.

Levitin, D. and Cook, P. (1996) Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 58(6), pp.927-935.

Apple. *Logic Pro*. (2009) (Version 9) [Computer program]

National Association of Music Merchants, (2014) *The 2014 NAMM Global Report*. [online] p.30. Available at: https://www.namm.org/files/ihdp-viewer/global-report-2014/A7352D4907B25A95B2CE27A075D3956F/2014MusicUSA_final.pdf [Accessed 6 Mar. 2015].

Perron, M. (1994)Checking tempo stability of MIDI sequencers. *Paper presented at the 97th Convention of the Audio Engineering Society, San Francisco.*

Peterson, P. and O'Dell, W. (1950) Selecting Sampling Methods in Commercial Research. *Journal of Marketing*, 15(2), pp.182.

Sacks, O. (2013) *Musicophilia*. New York: Vintage Books, pp.240-247.

Temperley, D. (2007) *Music and probability*. Cambridge, Mass.: MIT Press, pp.26.