KES Transactions on Innovation in Music: Vol 1 No 1 Special Edition - Innovation in Music 2013 : pp.67-78 : Paper im13bk-006

Wavefondler — a multi-touch interface for iPad to control audio on a host computer via a visualization of the waveform

Justin Paterson

N108, North Building, London College of Music | University of West London, St. Mary's Rd. Ealing, London, W5 5RF, UK justin.paterson@uwl.ac.uk

1. Abstract

It has long been a dream of those involved in audio manipulation to interact directly with a visualization of the target audio. In recent times, the mouse has been giving way to multi-touch interfaces, allowing a more tactile, immediate and intuitive interaction with the audio, and importantly offering more than one point of parametric contact. New modes of manipulation and performance are increasingly possible through a number of systems.

This paper will document the trajectory of interaction with audio visualizations and recent developments in such multi-touch applications, and investigate workflow and its implications via a number of case studies. Further, using a hybrid of emergent software platforms, it will demonstrate a custom interface design that allows the operator to access a visualization of an audio waveform on one or more iPads, and using multi-touch gestural control applied to the waveform, manipulate the sound via processing on a host computer.

2. Introduction

Multi-touch (MT) control is one of the most rapidly expanding areas of Human Computer Interaction (HCI). In audio production, particularly on iOS, there are a plethora of applications that facilitate various forms of mediation with visualizations of the audio stored and played back from the local device. Current devices tend towards categorization as sampler/synthesizer, DJ tool, or the novel.

Windows has featured MT support since version 7, yet few Digital Audio Workstations (DAW) on desktop and laptop machines are responding to this¹, despite the much stronger media propagation of this facility since the launch of Windows 8. Whilst there are a small number of systems available enabling MT via Windows, there is also the unique Slate Raven which features both dedicated

Copyright © 2014 Future Technology Press and the authors

¹ The MT facility of the OS cannot be used unless a given application also supports MT

hardware and software. In addition, there are multi-purpose touch screens such as CTOUCH which is multi-OS, but currently optimized (and only fully functional) for Windows.

Further to the above systems, there are a number of editor packages available for iOS, typically allowing creation of custom interfaces for MT control of Max/MSP using a range of knobs and sliders etc. Of these, only MMF-Fantastick [1] allows a visualization of the waveform with MT control whist running on Windows (only). The Cycling '74 Mira app allows realizations of many native Max/MSP objects on iOS, however at present it does not support the waveform~ object, which is the principal (amplitude/time) Max/MSP audio visualization tool.

The opportunity therefore exists to facilitate MT control of a waveform visualization on Mac-hosted audio; this is a primary function of the Wavefondler. The Wavefondler is a Max/MSP patch under MT control from Mira that recreates an image of the waveform, and allows the user to interact directly with this visualization, controlling audio on the host Mac computer, and when running in 'Max for Live' can effectively act as a real-time insert effect on any length of audio file, resident in Live, thus integrating into the greater DAW environment. To contextualize this, the Wavefondler also replicates a number of control features already available on commercial iOS apps (except again, this control is available for host-based audio), but in addition also allows a number of unique proprietary effects to be controlled.

3. Background and related work

3.1. Chronology

In order to contextualize the current state-of-the-art, a brief chronology of landmark products that influenced human interaction with visual representations of audio is now presented.

The representation of an image as sound, or indeed sound as an image is sometimes referred to as an audiovisual transformation. The earliest of these was created in the late 19th Century by André-Eugène Blondel with his invention of the paper-based oscillograph [2] which could offer a visualization of a telephone audio signal; the oscillograph was later to evolve into the more familiar oscilloscope.

Audiovisual transformation has been explored in numerous contexts ever since. Whereas the oscilloscope was of course purely representational, sound film was a largely unseen playback medium, however it was that which allowed the first creative generation of sound from shape. In 1930, Arseny Avraamov drew analogues of audio waveforms by hand before photo-reducing these for transfer to sound film. [3] Such approaches were notably developed by Daphne Oram — the Oramics system, which she developed for a number of years from 1957. [4] lannis Xenakis defined the UPIC 'syntax' in 1977, [5] which extended the earlier

haphazard approaches to visualization into a more defined system of timbres and manipulations.

Such work facilitated the generation of sound from image, and indeed empowered its relevance, but the true mediatory relationship between musician and audio visualization did not start until 1979 with the release of the Fairlight CMI. [6] This revolutionary and hugely expensive device featured a CRT monitor that could display a sampled or synthesized waveform that the user could interact with via a light-pen as seen in Fig. 1 (left).

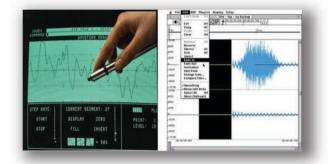


Figure 1. Left: The CMI light pen in operation; image from [7] Right: The Sound Designer sample editor GUI; image from [8]

By dragging over and reshaping the visual, the user learnt the audio reaction in terms of timbre and amplitude, and thus began a new paradigm in studio workflow.

In the 'mouse age', a further milestone was the Digidesign Sound Designer sample editor, released in 1985 and shown in Fig. 1 (right). This represented the first piece of software to display a visualization of audio that 'resided' on separate hardware (any of a number of supported samplers), and communicated via RS-422, a rather slow serial protocol. This software allowed fades and gain changes to be rendered, and provided visual feedback via the waveform image.

Hardware was quick to follow, and in 1988 the Akai S1000 sampler featured a small LCD screen with a scrollable split image of the waveform to view 'end & start' loop points more easily.

1990 ushered in the Opcode Studio Vision as shown in Fig. 2, the first MIDI plus audio sequencer. [9] For the first time, audio regions could be edited and placed on a time line alongside MIDI regions. This represented a major increment in interaction with the waveform as 'musical' placement and timing augmented the cropping and fading-type edits previously available in sample editors.

		Track 2: "recorder/		(D)
		MCI Helti: All	285 EM BARS	the second
1 Intelated	a to be to be to be to be to be to be	H H 10	11 12 18	in the second
3:		in the second se		
	The second and second and second and second			
81				
Audio-1				
Audia-2	ties ato - Ball and their lines	mar Marganger	better wetter	
_		and and have been as	1	
Augu-1				
Audie-4 book				
Audie-5				

Figure 2. Opcode Studio Vision; image from [8]

The audio waveform was now an iconic meme and developed in many products throughout the 1990s. Transient detection & slicing, time-stretching and off-line effects processing all became represented visually in what has become a familiar paradigm today and so will not be discussed further.

3.2. Multi-touch

Despite the principle first being realized in 1982 via an optical system [10], commercial (capacitive) implementation of MT is a recent development, and direct manipulation of a waveform is greatly enhanced by such interaction. Current MT devices fall into loose categories: Windows machines, 'iDevices'², non-iOS tablets and bespoke devices. Within the iDevice/tablet categories, there are both commercial applications and user-constructed interfaces constructed with editor software.

MT has been available on the Windows platform since version 7 in 2009. Although the OS supported this and MT-responsive monitors started to emerge, pro-audio manufacturers have been slow to implement MT. One reason for this could be the fashion for densely packed GUIs that do not lend themselves to finger control, and further, hands might often obscure important areas of the screen during operation. A full GUI redesign is not a trivial affair, especially given the enormous range of functionality that would need to be replicated on a larger 'finger-sized' scale. Released in late 2012, Cakewalk Sonar X2³ shown in Fig. 3 (left) is to date the only DAW to implement MT functionality. Whilst highly effective in offering movement of multiple faders simultaneously, the intricacies of audio editing have been avoided, and only pinch-zoom and scroll have been implemented with the waveform visualizations.

² This term is adopted as a collective to include MT Apple units such as the iPhone and iPad etc. Other tablet operating systems such as Android do not currently offer much audio-related functionality.

³ the current version is the X3



Figure 3. Left: Sonar X2⁴. Right: Stagelight's GUI⁵.

Whereas .NET is used for coding MT applications on Windows, JUCE is a C++ class library that can compile to several platforms: Windows 8, iOS and OSX etc. A device that is built with JUCE will respond to MT if run on a suitable platform, e.g. on Windows 8, the D16 Lush 101 plug in runs with MT capability in Cubase, despite the latter not being MT. [11]

In 2013, the Openlabs Stagelight of Fig. 3 (right) became the first of a new generation of DAWs that featured a GUI that was optimized for MT finger control, with larger and more spaced controls. It does not offer any interaction with a visualization of the audio.



Figure 4. Left: The Emulator⁶. Note the waveform displays, directly from Traktor underneath the 'overlay' GUI. Right: The Slate Raven MTX⁷.

An editor package exists for Windows; the 2010 SmithsonMartin Emulator, as shown in Fig. 4 (left). This device can host a custom MT interface that controls a third party application, and currently it seems to be closely aligned to Native Instruments Traktor — almost exclusively. In order to feature waveform

⁴ Image from http://www.youtube.com/watch?v=zh-fpA2-fto

⁵ Image from http://us.openlabs.com/2013/index.php/products/stagelight

⁶ Image from http://smithsonmartin.com/products/emulator-pro/

⁷ Image from http://www.slateproaudio.com/products/raven-mtx/

visualization, it allows regions to be defined in its 'overlay' GUI layer to allow the third party graphics to show through from beneath, but since Traktor does not support MT, these of course do not respond to it at present.

Perhaps due to Apple's legacy with music applications, iOS is the dominant contemporary platform for music-related apps. The taxonomy of theses apps tends towards devices which are holistic in their operation on the iDevice, act as 'fixed' functionality control surfaces for manipulation of an application on a host (desktop or laptop) computer which likely has its own control over DSP and audio functionality, or user-configurable editor packages that can control a target application on a host. Only the 'holistic' tend to feature audiovisual transformation.

Firstly, to discuss some examples of apps of the former 'holistic' type, Samplewiz (shown in Fig. 5 (left)) offers many classic synthesizer features and modes of playback. The most relevant here allows the user to polyphonically touch a visualization of a sampled waveform stored locally on the iDevice. Each finger can play a different (looped) section of the waveform, and the vertical axis transposes playback chromatically, although with a minimum resolution of ± 12 semitones, exact pitching can be hard to control. Despite this, it is a hugely gratifying and novel mode of performance; however, (although there is a two level zoom function) due to the fixed size window, if longer samples are imported, the relative size (on the x-axis) of meaningful sections of the waveform become too small to accurately highlight with fingers.



Figure 4. Left: Samplewiz, showing two-finger operation. Middle: Samplr⁸. The two white circles on the waveform represent where fingers touched. Right: Wavetable, showing puck-driven operation. Image from [12].

Another excellent example is Samplr of Fig. 5 (centre), which offers several modes of polyphonic playback, and a multi-track type facility. Playback modes include both grid & transient-based slicing and looping (that can arppegiate between selected slices) with volume on the vertical axis, a pseudo-tape mode where the virtual tape can be scrubbed in either direction with various enveloping and effects

⁸ Image from http://samplr.net/

functions. Again, the window size is fixed and as such the target audio is 'fitted' to the window.

Lastly notable here is Traktor DJ. Primarily a DJ tool, this app can run much longer pieces of audio then those above and features effective scrolling into manageable 'loop-windows', which can in turn be resized with two fingers, thus allowing its 'freeze' and 'slice' modes that play back individual samples from the longer audio.

The three apps above all feature the same degree of isolation from a host (DAW) and although featuring a degree of support for Core MIDI, typically require third party utilities such as Dropbox or Audioshare to import/export files to a host system; clearly an impediment to integrated workflow with a greater DAW environment. Although highly tactile and capable of delivering genuinely new modes of performance, they are also tied to purely 'live' performance, since there is little retrospective editibility of sequence/automation type events.

Apps with fixed functionality control include the Apple⁹ Logic Pro X Remote, which facilitates useful, but basic mixer functions.

There are a number of editor systems available for iDevices, tending to interact and control Max/MSP — C74, Lemur, Fantastick, and MMF-Fantastick (an extension of the former) all offer MT control, however of these only the MMF-Fantastick system of 2010 [1] offers a visualization when used on PC. This is important, since it was the first MT way of interacting with an audio visualization.

With regard to bespoke devices, in academic circles, the WaveTable [12] of Fig. 5 (right) appeared in 2008. This was based on the reactable framework, reacTIVision. [13] This system used a MT light table to display a visualization of a sample, and use Reactable-style hand-positioned pucks to manipulate parameters such as: zoom, selection, gain, LPF and erase. Although the system is visually arresting, it could be argued that the pucks might impede fluid real-time operation and accuracy. Further academic systems include the (DJ tool) Random Access Remix [14], which makes innovative use of dual timelines and a visualization of the waveform to orient the user. This system does not significantly exploit MT and is iPad native. There is currently one bespoke commercial system that facilitates MT on larger screens, released at Winter NAMM 2013 — the Steven Slate Raven interfaces [15] shown in Fig. 4 (right); only available for Pro Tools at the time of writing. Despite impressive MT control of the mix environment on a host computer with an number of GUI functions that optimize and speed workflow beyond native Pro Tools, these only offer single point control when editing waveforms.

⁹ It might also be noted that the new/revised GUIs in Logic Pro X are all 'finger sized', a strong hint towards Apple's desktop implementation becoming MT.

4. The Wavefondler

The Wavefondler Max/MSP patch was developed in order to provide a unique form of control. As will be noted from the preceding sections, it would appear that there is broad acceptance of the benefits of MT for performance and mixing operations, but it is clear that direct interaction with an audiovisual transformation on a desktop/laptop is yet to come of age, and that at the time of writing there is no way of using MT control on a visualization of audio that resides on an Apple Mac computer. The release of Mira in Summer 2013 offered new possibilities of implementing this. Mira is able to display a number of Max/MSP objects, however Mira does not include the waveform~ object, which is normally used within Max/MSP for audio visualization. As such, this did not represent a complete solution in itself, however Mathieu Chamagne produced an innovative Jitter-based abstraction (based on his MMF-Fantastick [1]) that represented the audio waveform using the multislider object, an object that can be displayed on the iPad using Mira. Rewire apart, Max/MSP generally functions as a standalone environment in normal usage; however, patches can be opened in the Max for Live (M4L) application within the Ableton Live DAW. M4L comes with a number of dedicated objects to extract timing from the host sequencer, and therefore can facilitate sample-accurate integration of the patch into a DAW environment with automatic tempo-matching.

In order to maintain a consistent magnification of the waveform visualization on the iPad that was easy to touch with fingers, it was decided to copy each progressing bar of real-time audio into a dedicated buffer for the real-time manipulation displaying only a single bar at a time; this incurs considerable latency at initial startup, but once the buffer is 'charged' then latency is minimal through continued use. The patch makes extensive use of the multitouch.mira object, which allows MT control over a prescribed screen area on the iPad. Each of the three separate tabs (shown in Fig. 5) provided an iPad screen that was divided up into various areas including some that were multi-function depending on which order fingers reached them from adjacent areas. The GUI comprised many layers in order to facilitate these objects and provide visual feedback.



Figure 5. The Wavefondler: the three different tabbed screen views.

4.1. Functionality and Operation

The Wavefondler waveform tab (Fig. 5, left) allows the user to specify areas of a bar for operation (in quantized note values of 1/4, 1/8/ 1/16 & 1/32) and provides three modes of playback: 1] slice playback with transport stopped, allowing the user to tap and play a selection, 2] overdub slice playback with transport running and output quantized to a clock, and 3] looped slice playback with transport running. Several effects processes can then be applied via MT: volume, filter, stutter, and combined filter & stutter. The user can select a slice for playback in the lower area, and then by sliding one or two fingers into the upper area, manipulate parameters associated with these effects, responding to the waveform visualization. Modulation was limited to 2-finger at the design stage, since earlier experiments with multiple fingers proved difficult for the user to meaningfully control. For volume, one finger controls volume itself whilst the other selects different slices. Broadly speaking, the stutter increases in rate as the finger moves higher on the y-axis, however there is an algorithmic process which plays back a variety of rhythms and pitch sweeps, tempo-locked in addition to 'just' finger placement. In the case of the filter, one finger's vertical position controls resonance and the other, cut-off frequency. Combined stutter/cut-off simply applies both at once. A matrix of eight presets is available to the user, and these can be freely switched between with a third finger. In addition, there are four horizontal lanes in the upper area of the waveform each of which allows one of: filter, stutter, chopper and transposer¹⁰ to be applied to a particular (quantized) area of each passing bar of the real-time audio¹¹ via 2-finger pinch-selection, independently of the real-time gestures described earlier.

The filter tab (Fig. 5, middle) features an x-y pad that allows 2-finger Independent control of cutoff and resonance of the LPF; this proved a very tactile approach. A second pad above allows control of the BPF alongside a frequency shifter. Again, there are a number of presets available on a small matrix that set numerous parameters on both filters. A one-bar envelope can be drawn directly on the waveform with MT allowing intuitive interaction with multiple waveform peaks. There is also a transpose facility that offers MT control over the pitch of 1/8-note slices of each bar, again easy to relate to the visualization of the audio.¹²

The chopper tab shown in Fig. 5 (right) uses a time-line that runs top to bottom, divided into note values as indicated in the coloured histogram running horizontally (in Fig. 5, there are 8 bins, thus representing 1/8-notes, although other note values are selectable). The right-hand edge of each bin aligns with points on the

¹⁰ See the chopper section below.

¹¹ This concept was originally a 'multi-track' development of the dBlue Glitch VST, however over the course of development of this patch, both 'Sugar Bytes Effectrix' and 'Illformed Glitch²' were released with similar (superior) functionality.

¹² Similar to the 'M4L Buffer Shuffler', which was unknown to the author at the time of development.

Justin Paterson

waveform, and it is a slice of the same note value at that point that will play from each bin. In Fig. 5, the displayed bar will play back 'as normal'. The user can adjust the histogram with MT to make different slices of the waveform play at different times, allowing a visual and tactile approach to beat-slicing. Different note-values can be selected to influence the slice-rate/size. In addition there is a transpose option on this page. There are two different-sounding algorithms, and the amount of transposition is selectable in real time via the piano-style keyboard. Transposition can happen simultaneously to beat-slicing.

In combination, the above feature-set offers a large number of sonic manipulations controlled by a relatively small number of gestures. The reader is strongly recommended to watch the explanatory demonstration video at: http://youtu.be/41B6FKAowXc

4.2. Multiple iPads

The Mira software allows the Wavefondler to function effectively using multiple iPads. Wavefondler was thoroughly tested using two devices simultaneously: a single iPad 2 and another iPad 3. The iPad 2 proved to have an insufficient data transfer rate over Wi-Fi to offer an functionally fast enough screen redraw, however music control information seemed prioritized, and was transferred with sufficient lack of latency for performance. Regardless, one problem that was encountered when using two iPads was the lack of proprioceptive control. When using a traditional hardware interface, the user can typically operate a fader or knob without looking at it, using a combination of proprioception and aural feedback to exert control. Since a touchscreen offers no such (hybrid) feedback at present and the eyes can only accurately guide the hands in a relatively small target area i.e. a single iPad, the use of two iPads did not literally give double the functionality of one. However, it was found possible to visually multiplex between the two quite rapidly, and in addition a limited degree of control was possible even with eyes averted from a given iPad.

Whilst there are systems under development which remove the need for such visually directed control (for example navigation devices that aid pointing to objects of interest with the user's own arm [16]) it is not conceivable at present how these could operate with sufficient accuracy for the type of control necessary in this context. Doubtless, suitable control will emerge in the future.

There were further issues around dexterity when using two iPads, since a twohanded approach had primarily been developed during software creation on a single iPad, however specific practice could negate this.

5. Conclusions

As can be seen in the demonstration video, the Wavefondler is a most intuitive and tactile device. It demonstrates novel interaction with the audio on a host computer, in the 'familiar comfort' of a DAW environment if preferred, doing away with the

Justin Paterson

tedious Internet-based file transfer often associated with other current audio apps, and crucially integrating with the greater multi-track environment. Although its sonic manipulation functionality is limited and at present oriented towards a glitch-type genre, it could be developed or modified to carry out many other functions through a similar interface and gestural control. These gestures are of course not novel, now becoming commonplace in a variety of apps, but here, they serve to demonstrate that they can be applied to host-based audio. Although not implemented in this version, it would be ideal to transfer all gestural information to a host sequencer for potential retrospective editing as automation.

At present, development and operation can be haphazard due to the formative evolution of Mira. That system does not currently pass data rapidly enough to consistently transfer complex graphics (such as the multislider object or textual elements) in real time, and also suffers memory problems and these can impede workflow. Doubtless these problems will be solved in a future revision of this potentially wonderful software, and as the Max/MSP community embraces it, a plethora of devices will emerge.

The question arises of how long such an approach can hold currency. Windows has dominated the desktop and laptop MT market for some time, and Apple will not hold back its response forever. It is quite likely that their response will be a step increment beyond 'mere' MT, potentially incorporating (further) voice, 3-D gesture, and retinal-tracking functionality. Music software manufacturers have been slow to develop GUIs of existing products optimized for MT. This task is daunting, and it is conceivable that they are waiting to see the future direction of Apple desktop machines before committing, but it will come. Regardless of how Apple eventually implements MT or its successor, the remote iPad control of the waveform will likely become redundant when the user can interact directly with the host screen. As such, the Wavefondler can be regarded as a novel yet interim device.

6. References

- M. Chamagne, "MMF-Fantastick | Mathieu Chamagne." [Online]. Available: http://www.mathieuchamagne.com/2010/08/25/mmf-fantastick/. [Accessed: 02-Dec-2013].
- [2] J. Miguel Dias Pereira, "The History and Technology of Oscilloscopes," *IEEE Instrum. Meas. Mag.*, Dec. 2006.
- [3] J. Thoben, "Technical Sound-Image Transformations." [Online]. Available: http://www.see-this-sound.at/print/51. [Accessed: 13-Dec-2013].
- [4] P. Manning, "The Influence of Recording Technologies on the Early Development of Electroacoustic Music," JSTOR Leonardo Music J. Vol 13 2003 Pp 5-10, 2003.
- [5] G. Marino, M.-H. Serra, and J.-M. Raczinski, "The UPIC System: Origins and Innovations," JSTOR Perspect. New Music Vol 31 No 1 Winter 1993 Pp 258-269, vol. 33, no. 1, 1993.
- [6] "Fairlight The Whole Story." [Online]. Available: http://anerd.com/fairlight/fairlightstory.htm. [Accessed: 02-Dec-2013].

- [7] "FairlightUS Directory Listing of /downloads/Photo Albums/30 Years of Fairlight Captured/." [Online]. Available: http://www.fairlightus.com/downloads/index.php?dir=Photo+Albums%2F30+Ye ars+of+Fairlight+Captured, [Accessed: 12-Dec-2013].
- [8] C. Halaby, "KVR: 'It was 21 years ago today...' How The First Software DAW Came About."[Online]. Available: http://www.kvraudio.com/focus/it_was_21_years_ago_today_how_the_first_so ftware_daw_came_about_15898. [Accessed: 02-Dec-2013].
- [9] "1990 Opcode Studio Vision Digital Audio Sequencer | Mix Inducts Opcode Sequencer Into the TECnology Hall of Fame." [Online]. Available: http://mixonline.com/TECnology-Hall-of-Fame/1990_opcode_sequencer/. [Accessed: 12-Dec-2013].
- [10] N. Mehta, "A Flexible Machine Interface," M.A.Sc. Thesis, University of Toronto, Tontoto, Canada, 1982.
- [11] R. Vincent, "Windows 8 Multi-Touch in Music Production Part 2 (of 2) with Sonar, Cubase, Live, Reason, Emulator - YouTube," 2013. [Online]. Available: https://www.youtube.com/watch?v=4G4jlpdTuhY. [Accessed: 29-Nov-2013].
- [12] G. Roma and A. Xambo, "A tabletop waveform editor for live performance," in *Proceedings of the 8th International Conference on New Interfaces for Musical Expression*, Genoa, Italy, 2008.
- [13] M. Kaltenbrunner and R. Bencina, "reacTIVision," in *Proceedings of the 1st international conference on Tangible and embedded interaction*, New York, NY, USA, 2007, pp. 69–74.
- [14] J. Forsyth, A. Glennon, and J. P. Bello, "Random Access Remixing on the iPad," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, 2011.
- [15] "RAVEN MTX: Slate Pro Audio." [Online]. Available: http://www.slateproaudio.com/products/raven-mtx/. [Accessed: 13-Dec-2013].
- [16] T. Ahmaniemi and V. Lantz, "Augmented reality target finding based on tactile cues," in *Proceedings of ICMI-MLMI 2009*, Cambridge, Ma. USA, 2009, pp. 335–342.