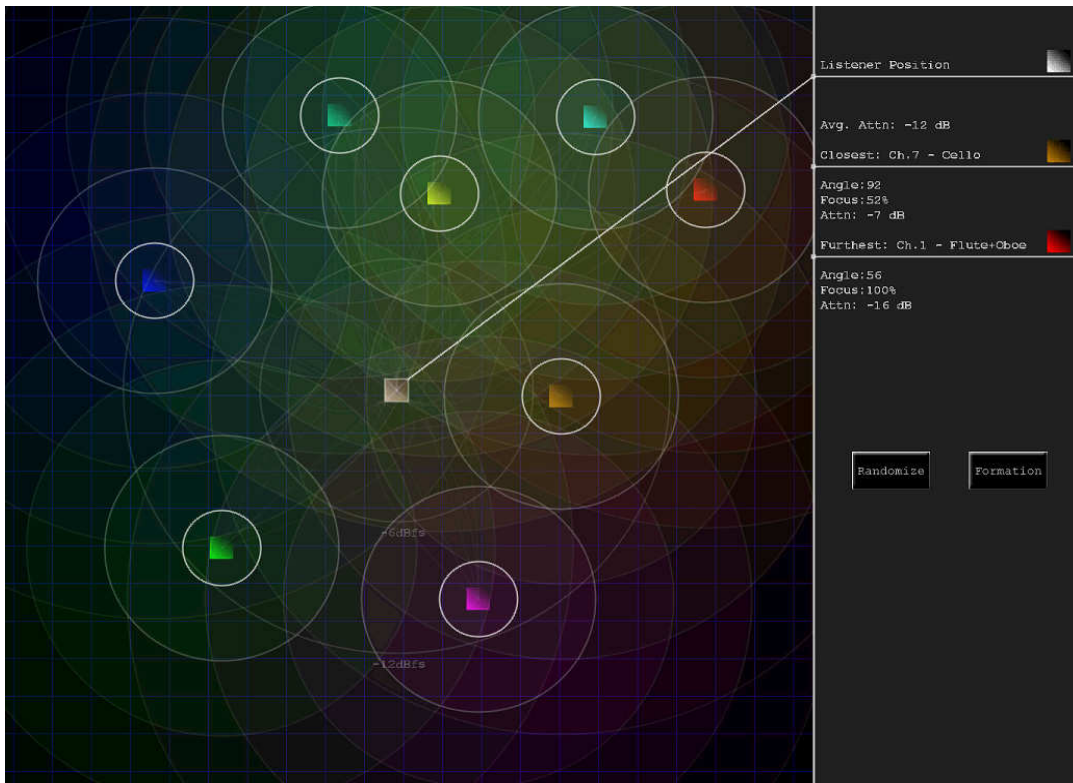


AWOL: Control Surfaces and Visualization for Surround Creation

Vincent Diamante

Interactive Media Division - University of Southern California
2276 South Figueroa Street
Los Angeles, CA 90007
213-840-0645
diamante@usc.edu



Abstract

This paper describes the creation of a surround sound manipulation tool and the development of pieces of sound and music where this tool, rather than a mixer or traditional music input instrument, is at the center of the piece's production. Using a tablet PC connected to a traditional digital audio workstation, surround sound mixing data is manipulated in a dedicated high precision, high feedback environment, better enabling sound spatiality to comprise the main musical structure of the piece. Both the possible end workflows and the resultant pieces offer alternatives to computer audio content workflow as inspired by traditional analog studio mixing console paradigms.

Keywords

Tablet PC, Surround Sound, Surround Mixing, Control Surface, Touch Screen, Pen Computing, DSP, Sound Visualization, MIDI, Mixer, Plugin, Aleatoric Music

Introduction

The development of music and audio content has been strongly influenced by the technology surrounding it. As technology is pioneered for the performance of new types of sound (effects, DSP, multi-channel, high definition), sound content creators show quickness in experimenting with and adopting new tools and workflows. A thriving marketplace for professional and consumer musical instruments proves the space for technological growth.

Surround sound in all of its permutations has enjoyed widespread adoption by consumers, prompting producers and developers to create increasing amounts of surround content. This is especially true in movies, where soundtracks over the years have required progressively more and more channels of audio, from matrixed 4 channels to 10+ discrete channels of audio, each channel sounding via its own individual speaker or group of speakers. As channel count goes up and content grows more sophisticated

in its composition, the audio experience becomes potentially more arresting.

Surprisingly, tools specifically for surround sound creation are few in number and minimal in their presence. Integrated surround functionality is offered in the traditional mixer as a joystick taking the place of the pan pot in typical channel strip topography. Software plugins mirror this mixer functionality on the digital audio workstation, but are similarly restricted to the individual channel strip. Other plugins and effects processors are available to do simulated surround mixing of premixed content through phase processing. Still other pieces of gear are processors that take multiple single channel and stereo sound files in and output a proper surround format audio stream.

These tools require that most of the work needed to manipulate sound attributes like perceived physical size or distance actually occurs elsewhere in the audio workflow, outside of the limited surround sound specific tools provided.

AWOL seeks to rectify that by supplementing the traditional workflow of a sparsely engaged, multiple control channel topography with an overloaded control surface/visualization tool which hooks into multiple sound attributes and displays the states of multiple sounds relative to each other, affording the content creator a picture of the surround sound space as a whole.

“AWOL” is a reference to the US military status Absent Without Official Leave. Like a well commanded military brigade, audio engineering workflow can be extremely regimented, and typically “good” audio engineering workflow is a well-regimented procedure, with veteran engineers, trade schools, and universities teaching hard sequences that form the backbone of content creation practices. As amateurs, hobbyists, and general consumers receive more ability to create and distribute their own works, the tools targeted specifically at these markets have found their way into all spaces, subverting the idea that a specific audio creation workflow is an ideal workflow. AWOL is designed to offer freedom away from traditional mixers and channel strip topography to its users as they create surround sound spaces.

Prior Art and Observations

Project AWOL draws inspiration from the surround sound technologies that preceded it, as well as from various examples of surround sound content.

Disney’s Fantasia in 1940 was a strong example of content driving technology. In the course of developing the first multichannel surround sound film, Disney engineers developed techniques that are now commonly used in audio engineering, including click tracks and overdub recording. Most importantly to surround sound was the development of a device known as “The Panpot” allowing fading a sound

between two different speakers to mirror the motion of a source from one end of a theater screen to the other.

Since that time, the technology has become more ubiquitous, though not significantly more sophisticated. Disney’s Fantasia was a spectacle installed in only three cities in America at once, while thousands of movie theaters around the world are able to playback surround soundtracks. The difference in delivery format is clear, with modern surround sound having more channels (six to eight discrete speakers) and more specialized channels (for LFE low frequency speaker and speech for the center channel). The technology to create this content is not far removed from the 1940s panpot. Instead of a panpot moving a single sound source between two speakers, the modern audio workstation or mixing console features a joystick moving sound on a two dimensional plane, such that closeness to a speaker on that plane feeds more of that sound source to that speaker than to other speakers in the surround sound arrangement. Other configurations have axes on the joystick essentially mirroring the functionality of a pair of one dimensional panpots.



Figure 3 – Surround Effects Processor for Quad Sound (QUEG)

In the digital audio workstation software world, plugins mirror the functionality of these hardware joysticks. Instead of a physical joystick, an icon is moved in a rectangular or elliptical space to indicate to which speaker or speakers represented in that space the sound is biased.

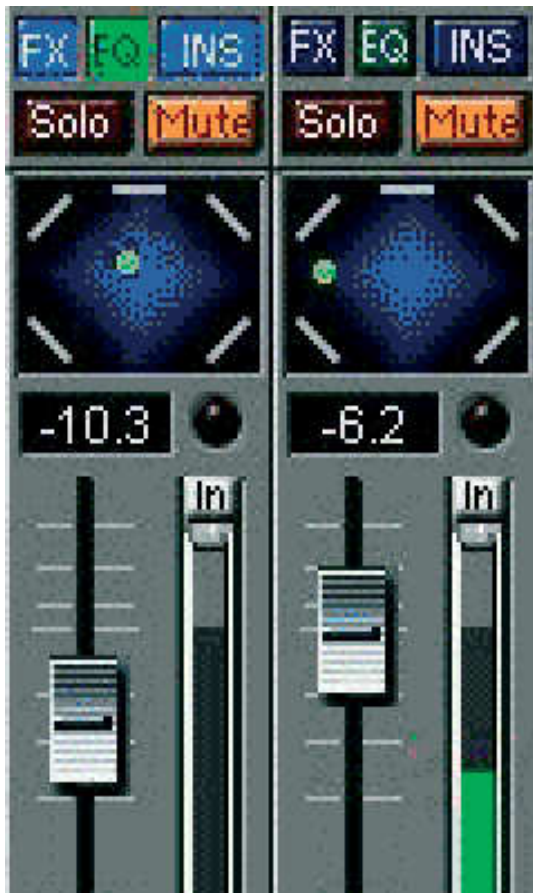


Figure 4 – Surround Panning on Nuendo Channel Strip (Steinberg)

Chief among the drawbacks of this model is the inability to visualize the space as a whole. On a single channel, it is easy to see where the sound is in apparent existence relative to the physical space's speakers and the intended listener position. As channel count goes up, it becomes extremely difficult to keep track of the surround position of each source. It is even more difficult to keep track of the surround position of each source relative to each other. With modern pop song and movie soundtracks easily requiring dozens of tracks, accurately surveying the surround placement of individual channels within the mix become difficult through traditional channel strip topography in both the physical and the virtual worlds.

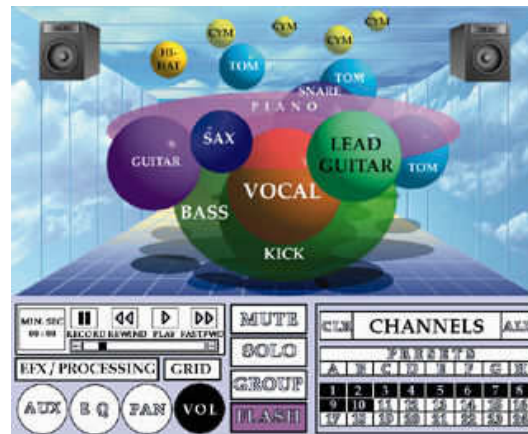


Figure 5 – The Virtual Mixer places sound sources as 3d spheres placed in front of or behind other sources (Gibson)

David Gibson's Virtual Mixer visualizes sound sources as spheres existing in 3D space. In its standard stereo mixing configuration, spheres that are in the foreground are most prominent in the mix, with a higher channel volume. Attenuating the source moves it into the background. This allows the engineer to see how certain sounds can mask other sounds in the complete mix. The Virtual Mixer also provides a surround sound mix mode. In this mode, front to back relationships no longer correlate with volume but with a more traditional front speaker to rear surround speaker differential.

While the Virtual Mixer makes it much easier to visualize placement in a complex mix as compared to traditional channel strip topography, spatiality itself is not directly addressed. While a sound source's corresponding sphere can be dragged to the edge of the Virtual Mixer's 3D space, the sound itself is not appropriately rendered as volume is no longer correlated with distance. This is a trait shared with traditional joystick point style surround panners as well.

In the video game development world, surround sound has become almost mandatory in the creation of immersive 3D games. With the proliferation of high powered audio engines, surround sound implementation has become relatively easy as surround sound is correlated directly with object position in 3D space in real time. The resultant surround effect in the final game is very effective. It is notable that with most video game soundtracks, spatialized sound effects are most felt as the player and his viewpoint moves through a space of static sound elements, while traditional audio engineering moves sound around a central listener point. Also, these game audio tools greatly lack the flexibility of a traditional audio editing environment, not designed to interface with digital audio workstations. The internal game audio development tools are often lacking in robustness and granularity of editing ability provided to the game audio designer. In the popular Unreal engine, the designer has little input other than to

correlate an Unreal actor (object) with a sound cue, a volume, and an attenuation radius.

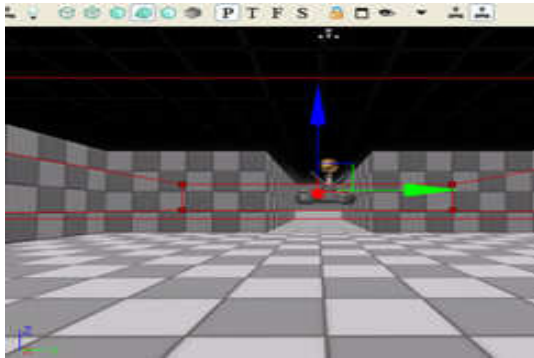


Figure 6 – UnrealEd allows sounds to be placed precisely in 3D space, on both static and dynamic actors (Epic)

In being inspired by this disparate array of tools, Project AWOL seeks a place that combines the precision afforded by modern high powered 3D game engines and the flexibility of traditional audio engineering.

Project Description

AWOL is designed to ameliorate several difficulties seen in traditional audio development paradigms. Console mixer topography is a strong obstacle to the perception of the surround mix as a whole. The use of multiple joysticks or plugins to affect a large number of sound sources makes manipulation difficult. This difficulty increases as the sound channel number increases. Also, the distance of a joystick's throw or a sound icon's displacement from a center point is never representative of the apparent distance heard. Precision far beyond physical joystick manipulation is also desired.

From this, AWOL is designed to be fast to manipulate, allowing rapid iteration of sound mixes with large changes to surround data and immediate, easily understood feedback of those changes visually. The AWOL tool should also manipulate more than speaker mix ratios, attenuating channels and frequency bands of sound sources to make the visual distance shown by the tool representative of the perceived distance of the sound over the speakers in a mix.

From the beginning, the project was intended as a piece of software for the Tablet PC. In addition to the screen for visual feedback that a dedicated Tablet PC provides, high-end Tablet PCs are outfitted with active digitizers provided by Wacom, such that pen location both on and above the screen surface is detected through electromagnetic resistance. This method affords extremely high precision over standard touchscreen passive digitizers (with finer resolution than the screen itself), and more possibilities for inputs, including hovering and pen angle. The pen also adds far greater quickness in manipulating

disparate items on the screen in a precise manner, as compared to a standard mouse.

Development of AWOL was through processing and the promidi 2.0 processing library. Promidi is used as the primary communication protocol, potentially allowing AWOL to take over the duties of other MIDI-based control surfaces, mixers, and audio workstation software.

The primary development system consisted of two computers. One computer is the main workstation, running Cakewalk's Sonar 5 Producer, a popular DAW software solution with surround provisions. The second computer is the Tablet PC, running a processing app, which interfaces with Sonar 5's capabilities through mapping of MIDI continuous controller messages to remote hooks. These remote hooks allow the Tablet PC to appear as a generic MIDI control surface to manipulate the following: Volume Fader, Surround Angle, Surround Focus, Surround Width, Hi Shelving EQ.



Figure 7 – AWOL, displaying information on channel attributes

The Tablet PC screen shows the position of each sound channel in a Sonar 5 project as a colored box positioned relative to a listener point (depicted as the center of concentric circles). These points can be placed and dragged throughout the screen, and changes in a position is immediately reflected in the sound through constant MIDI CC data transmit.

If a point is placed relatively close to the listener position, the surround width becomes slightly wide and focus becomes lower. While the sound is coming to the listener at an angle, this is less perceived than the fact that the listener is almost on top of the sound. As the point is dragged slowly away from the listener position, focus becomes larger and width becomes narrower, making the sound's directionality more apparent. As the point is dragged still further, focus continues to grow, width becomes still narrower, and the sound just begins to attenuate. When dragged still further away, EQ begins to affect the sound, shelving the high frequencies 7 kHz with a low Q.

This manipulation of multiple attributes simultaneously as one moves the "position" of one sound creates a paradigm that can be viewed as the exact opposite of the traditional mixer model. Rather than manipulate multiple, diverse, and disparate attributes (through faders, pots, and joysticks on a mixer) in order to achieve an effect, AWOL has the user immediately

achieving the effect while the underlying program deals with all of the attributes, hiding the user from those attribute values as well as the math used to arrive at those values.

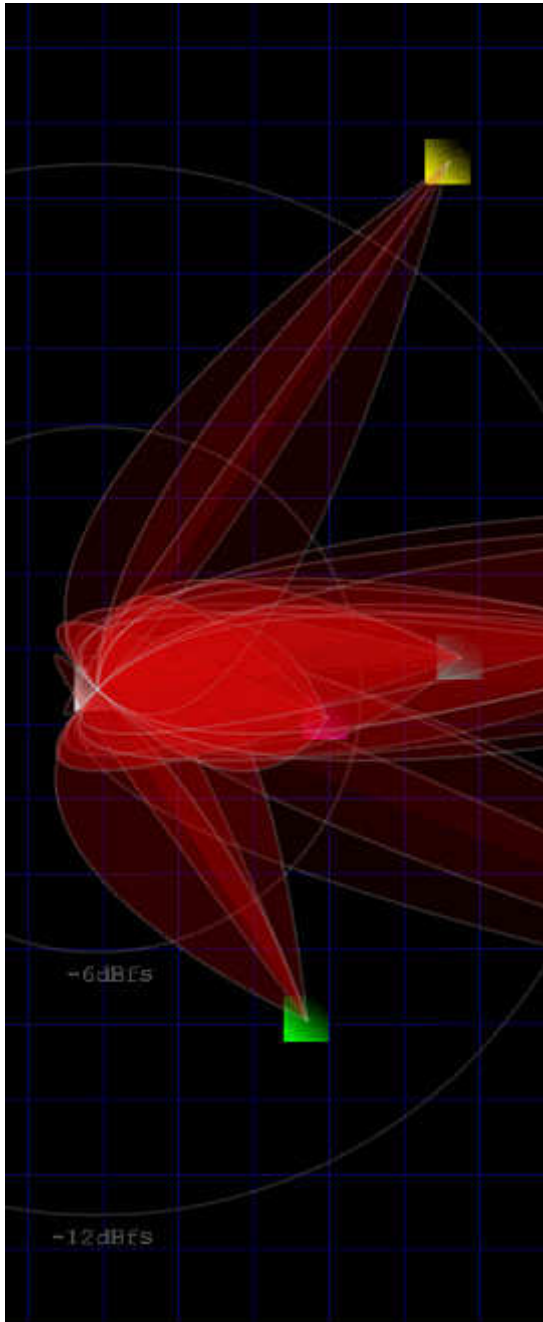


Figure 8 – AWOL, displaying sound points as colored squares and the listener position a white square being barraged by red sound

A feature not found in the above mentioned surround sound mixing methods but readily apparent in surround sound for video games is the ability to manipulate the listener's position. In a video game, the listening position corresponds either to the camera view or to an on-screen avatar. In AWOL, the

listener is a position that is manipulated just as the channel sounds are manipulated.

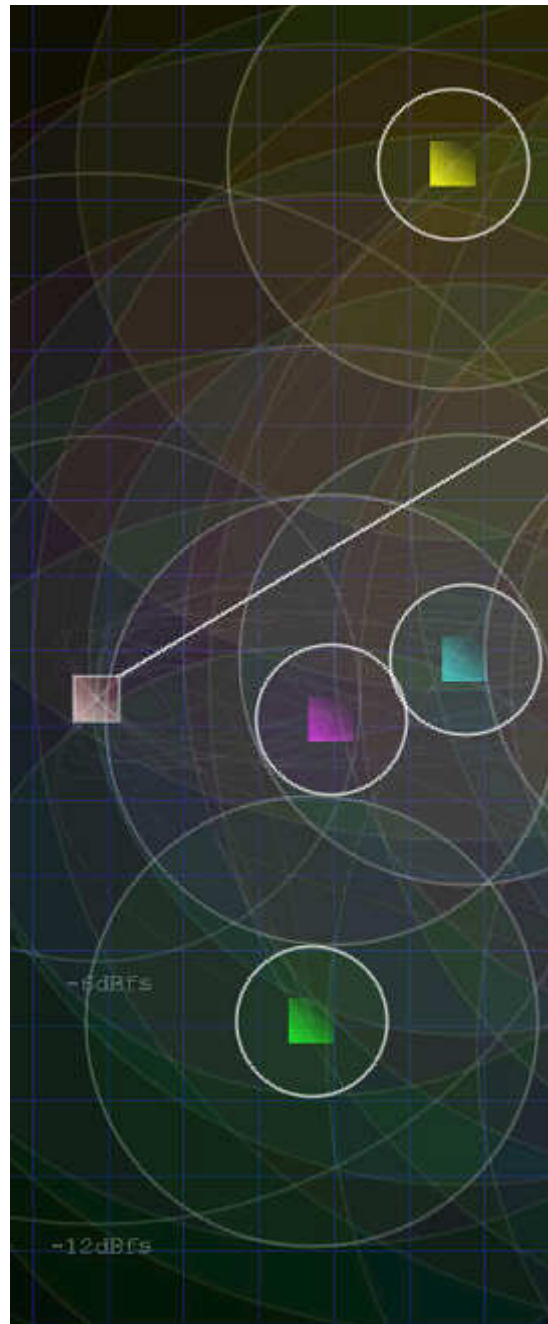


Figure 9 – AWOL in listener mode, displaying sound points as colored squares radiating their color filled concentric circles and additive blending between sound points. The user can place the listener position by quickly seeing the intensity levels at positions within the sound space

As the listener position is dragged around the screen, AWOL changes its display mode, communicating the sound sources as points within concentric circles. Under this model, the user immediately sees where sound intensity is greatest and least in the sound arrangement based on the color intensity. Though

there is no listener position in MIDI or audio sequencer parlance, Project AWOL communicates this by adjusting the attributes of all the sound channels simultaneously as the user drags the listener position across the screen. For an 8 channel audio mix, this means that 40 continuous controller messages are sent every update when manipulating the listener position.

With the pen as the input device, it becomes almost second nature to select multiple points and place them very precisely in specific spots in sequence. Forming and visualizing groups of sound sources is extremely easy. The speed with which sounds can be manipulated allows for extremely fast shifting between completely disparate sound setups.

A single dedicated screen to surround sound visualization makes it extremely easy to see a visual shape to the mix as a whole. The square icons make it simple to create groupings of specific sound sources and move them precisely around the sound space. It is also easy to see possible holes in a surround mix, even before listening to the sound.

With this visualization also comes the ability to preconceive the music composition, audio recording, or remix as a specific visual shape. This opens up new possibilities to the description of sound. Visual shapes are easily labeled and described among all people, while audio setups and sound placement can be nebulous when speaking strictly in terms of audio or individual channel configurations. The concept of lines, triangles, or curves of would be difficult to visualize and more difficult to execute when seeing and manipulating sound channels one at a time as with a channel strip based mixer. More difficult shapes and formations would be nearly impossible to visualize with a moderately-sized to large mix.

The ability to manipulate the listener's position also opens up the concept of exploration in the audio mix. With the sound sources set up as seen fit, one may explore the space by experimenting with the listener's position. With the automation of the shifting of all channel's surround data at once, the resultant composition of the soundscape verges on aleatoric. The interaction is almost like a video game in execution; just as a player might explore a 3D world, manipulating an avatar to find a hidden treasure, the audio engineer or even the music composer may seek to find an ideal perspective in a sound space.

Production Examples

Pop Music

During its development, AWOL was used to aid in the surround remixing of multichannel music tracks by Chris Oroc. Here, AWOL was used as a generic MIDI control surface interfacing with a Pro Tools 6 DAW.

As either abstract collections of ostensibly human performed components or human performed pieces of music with few abstract elements, it was decided that there was little need for automation of surround data, or sound position changing over the course of a given track. The remix sessions were dominated by discussion of instrument placement and the shape of the surround space. The discussion ended up sounding more like a choreography session than a recording session, as people spoke of placing cracking whips in the audience and having a big band take up the whole width of a theater stage.

These remixing sessions moved extremely quickly from song to song, especially as songs on later sessions had musicians coming to the song with an idea for a sound formation which would be executed before previewing the song in surround sound, which would then force, at most, mild tweaks to individual sound positions.

Visual Novel (Game) Soundtrack

In the traditional paradigm of visual novel video games, static images are the only visual indicators of environment and, as such, rarely are strong indicators of space definition. As this production follows this model for visuals, it was decided early on in production to use surround sound (a rarity for this game genre) to suggest and define space so that the player could focus on text rather than visuals.

Production continues on the sound for this game, and early experiments indicate that the AWOL tool is extremely appropriate for this kind of sound creation. Coupled with sequencer automation, AWOL is extremely adept at taking stock sound library material, especially those for environments and people, and placing/moving them in relation to the listener's space for maximum realistic effect. By setting the attenuation factors on AWOL much higher, it was possible to do all moves through real-time AWOL control rather than creating envelopes. With the right numbers in place, a conversation coming into earshot, loitering around the space, then passing out of audible range is doable in one take of automation recording.

While this sort of sound environment design is in the context of game creation, the above would be equally applicable in creating surround sound content for video.

Multi-camera Live Venue Recording and Spot Micing

AWOL was not used in a live venue recording, but the simplicity of creating surround mixes of audio from different perspectives suggests that it would be quite adept in this. For a straightforward surround mix of a play, one could simply create an arrangement of sound channels that would mimic the arrangement of microphones in a venue: stage mics in the front,

crowd ambience mics in the rear, and lavaliers that move through the space via sequencer MIDI automation. If a recording of the event at that venue was from multiple camera angles, setting up mixes for each of the different perspectives would be simply moving AWOL's listener position to correlate with camera position relative to microphone position. This presents significant time savings to the audio engineer should multiple surround sound mixes for multiple video streams be required.

Original Orchestral Music

During the production of Project AWOL, I wrote several minutes of electronic mockup of orchestral music in atonal/neo-tonal style. Though composition began with a fully-functional AWOL tool at hand, I chose to leave surround mixing to the very end of the process.

As an experiment, I decided to write the music by purposely pushing the musical sound of the composition to what would be considered a "muddy" or unreasonably thick sound when monitored in stereo. I used many techniques that would be considered traditionally bad in orchestration: writing instruments in weak parts of their sounding range, creating clusters of instruments sounding continuously in less than an octave's range, and having flourishes and otherwise sharp ornamentation by naturally weaker instruments lost in sustains from stronger instruments.

Once the orchestral suite was completed, I mixed the music with Project AWOL. Rather than mix with particular abstract shapes or standard orchestral arrangements in mind, I used AWOL's combination of attenuation and apparent sound position control to thin and create space for the sound to breath. Arrangements that would have been unacceptable in a two channel listening environment were made not just acceptable but enjoyable with a multi-channel monitoring setup and the ability to tweak to that setup. The experiment proved fruitful as the music was well received while the amount of time required to get to final mix for six minutes of music was surprisingly low at less than half an hour. During that time, I was able to immediately hear multiple viable surround mix setups through both the general speed afforded by the tablet PC platform as well as the ability to create random arrangements of sound. These random arrangements could be had with a quick pen tap and provided excellent starting points for surround formation tweaking.

Limitations

While AWOL has seen some success in its limited use, there are strong limitations on the current incarnation of the tool.

The most frustrating limitation is the fact that it is a MIDI control surface. As such, it is limited by MIDI

bandwidth. As channel usage goes up, so too does the bandwidth needed as data for five double byte continuous controllers are sent every update. With 30 channels and updates mirroring half NTSC speed (30 Hz), the MIDI transport is nearly saturated by 30 kilobytes of data per second. With the automation data sent back from the sequencer, it is possible to exceed the bandwidth of MIDI from Sonar to AWOL, as Sonar does not have good management for cleaning MIDI output of extraneous data.

Under the current graphical model, the AWOL processing app is extremely irregular in its screen update. Above 25 channels, the visual framerate may update between 5 and 10 frames per second, which poorly represents the actual surround data being sent to the host sequencer.

Currently the input is extremely loaded, as three axes of data (x position, y position, and rotation) are factored to create MIDI CC data for 6 outputs. Decorrelating on an sound channel level requires changing hard coded content in the source, and this should be moved to the AWOL instrument surface level before it becomes a useful tool.

Ideally project AWOL would function as a USB device that hooks in to VST/DXi/RTAS plugins on all channels, allowing full sequencer native resolution handling of all attributes, rather than 16 bit handling.

On a related note, AWOL would lose little functionality if existing only as a virtual window surface within a typical DAW's audio sequencer. One could argue that some speed is lost moving from the tablet PC model to the desktop paradigm; however, the tighter integration that is possible with a software-only plugin product rather than a software-hardware hybrid could make up for this.

Acknowledgements

Thanks goes to Perry Hoberman, Mark Bolas, Scott Selfon, Peggy Weil, Michael Naimark, and all other faculty and advisors who have helped me on this project. Special thanks goes to Chris Oroc for providing sample music to work from. Special thanks also goes to Dr. Seung Park for his input on game soundtrack development.

References

- [1] Epic Games, *Unreal Developer Network*. Retrieved 10 February 2007. <http://udn.epicgames.com/Main/WebHome.htm>
- [2] Gibson, David and Peterson, George. (1997) *The Art of Mixing*. ArtistPro Press.
- [3] Gough, R. Hunter. *In The Pit*. Retrieved 10 February 2007 <http://www.studiohunty.com/itp/>

- [4] *HUI*. (1997) Hardware input device, United States: Mackie.
- [5] Kay, Johnathan. *Film Sound History*. Retrieved 10 December 2006.
<http://www.mtsu.edu/~smpte/timeline.html>
- [6] *Lemur*. (2006) Hardware input device, France: JazzMutant.
- [7] *Nuendo*. (2002) CD-ROM, Germany: Steinberg Media Technologies.
- [8] Nunan, Michael. "Surround Sound - Now and in the Future" 121st AES Convention. 6 October 2006.
- [9] Nyman, Michael and Eno, Kenji. *Real Sound: Kaze no Riguretto*. Warp Productions, 1997.
- [10] Orr, Tim. (1975) *Quadrophonic Effects Generator*, EMS Music Systems.
- [11] *Pro Tools 6*. (2006) DVD-ROM, United States: Digidesign.
- [12] Schmitt, Al. (2004) *Genius Loves Company*. Concord Records.
- [13] *Sonar 5*. (2005) DVD-ROM, United States: Twelve Tone Systems.
- [14] Young, Rob. (1996) *The MIDI Files*. Prentice Hall