or any Appendices.

Enhanced Decorrelated Audio Upmixing Through Spatialization

Joshua Dickinson
Institute for Clarity in Documentation
1932 Wallamaloo Lane
Wallamaloo, New Zealand
trovato@corporation.com

Ryan McGee
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Matt Wright
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

ABSTRACT

Traditional chorusing and source multiplication techniques tend to suffer from phase cancellation at the micro level and unnatural delay-time predictability at the macro level, both of which subtract from the illusion of multiple sources. These limitations often make it difficult or impossible to simulate natural, dense, soundscape from sparse input material. Quite often, stemming from either practical limitations or the desire to create very specialized sonic scenes, a recording artist or composer wishes to make one source sound like many. Delay line multiplication would be an ideal solution were it not for these inherent difficulties.

A piece of software is introduced which utilizes spatialization and custom 'ensemble-effect' algorithms in order to create a more 'realistic' multiplication of input material. The result is a user-friendly yet highly customizable tool, suggesting many practical applications for music and sound design. The creation and result of this program is discussed while being illustrated with many sound examples[1].

Keywords

Spatialization, chorus, upmixing, ensemble effect, Sound Element Spatializer, multiplication

1. INTRODUCTION

Following the work of Mark Dolson[1], there have been attempts to achieve the 'ensemble-effect' through small randomized modulations in the amplitudes of harmonics. The most notable study in this regard was that of Daniel Kahlin who compared the synthesized ensemble to that of a real ensemble in the spectral domain in order to understand the differences between each. He eventually concludes that:

> "Pure amplitude modulations alone cannot simulate all aspects of the ensemble sound. The reason is that the notion of a signal amplitude is tied to a time window of measurement that is too long to capture the spectrum smearing properties of an ensemble at high frequencies." [2]

[1]These can be heard at this semi-permanent link: http://mat.ucsb.edu/dickinson/

NIME'11, 30 May–1 June 2011, Oslo, Norway.

Earlier attempts to create the illusions of multiple sources tended to focus on small time delays of identical tracks, however this usually generates unnatural phase cancellation or flanging. Known as the Chorus Effect, a particular implementation of this formula has become an extremely popular device available in most commercial DAWs. While it does somewhat create the impression of an ensemble from a single source, it is more of a super-natural rather than natural effect. It increases the size or presence of a single source, however few people would confuse this sound for a real group of instruments, singers, or other sounding objects. Additionally, parameters are usually extremely limited, and do not allow for the subtle control that a user would ideally want if he/she were trying to recreate a natural sonic scene.

I introduce here a piece of software which achieves the realistic multiplication of audio sources through three dimensional spatialization, filtering, and a novel delay-line based approach. It enables the user to create a seemingly unlimited number of variations through the creation of an arbitrary number of duplications of a single sound source, as well as the ability to place each of these duplications in their own time/space locations and velocities.

2. OVERVIEW

"Source Multiplier" is a program used to create a number of effects, based upon the duplication and time-varied delay of one or more input recordings. The user first selects the number of duplications to create for each track and then positions each of them in their own unique location in space and time. Specifically, parameters are set to determine the amount of 'base delay' or fixed distance from real time (base delay = 0), 'delay waver' or size of the space that the signal is free to move around, position in 3D space, general spatial velocity, and type of spatial movement. Through these five basic parameters, one is able to create many variations of distinct sonic scenes. While essentially infinite, these can be generalized into a few basic categories in terms of their perceived effect and function:

1. Natural
   - Ensemble
   - Swarm
   - Unrelated Sources
2. Unnatural
   - Flanging or Chorus
   - Warped Pitch

These effects are by no means discrete and separated-indeed they often blend with one another, however they do represent the basic range of possibilities of this process. Their classification is additionally complicated due to the fact that they are determined not only on the parameters chosen by the user, but are also largely dependent upon the type and quality of the original input signal. Any additional reverb or effects on the signal, or the inclusion of multiple sources (polyphony) in the original recording will change the perceived effect significantly. The cause of this will be discussed in the following sections, however it can be summarized in the fact that any sound in the original track is shifted and moved in unison which is very recognizable sound that is rarely if ever heard in the natural world.

3. TECHNICAL

This program is written in C++ using the Gamma library written by Lance Putnam. Spatialization is done in Sound Element Spatializer (SES) written by Ryan McGee.

> "It seems likely that we would obtain progressively better ensemble simulations if we were to expand on the time-delay approach, with more numerous signal replicas and better models for the delay functions."[2]

In building this program, I chose to replicate as much as possible the natural phenomena of multiple instruments or sounding objects being heard in a real environment. If we take the example of multiple instruments playing a unison line in an ensemble, there are a number of factors that give this sound its character:

1. Each instrumentalist is slightly out of time with each other on the millisecond scale, and this discrepancy of timing is not constant or fixed.
2. Similarly, each instrumentalist is always slightly out of tune with one another on the scale of a fraction of a semitone.
3. Each instrument has unique acoustic properties (spectral identity).
4. Additionally, each instrument possesses its own position in physical space which in turn determines its volume and amount of fixed delay relative to a stationary listener.

4. TIME

To simulate the first phenomenon, each duplicated voice is passed through a delay line that moves within a fixed area around a predetermined 'base' or anchor time. This successfully creates the impression of multiple sources, however there are a number of unwanted artifacts that arise from this process. Our ears are extremely adept at picking out patterns in sound, which means that each delay-line cannot simply move back and forth through an LFO, as this creates an easily discernible repeating sequence. Neither can this delay amount skip about randomly, as any sudden shift is perceived as an unnatural break in continuity. An intuitive solution would be to ditch the time-varying delay line entirely, in favor of a fixed one, however as discussed earlier, this model causes unnatural delay-time predictability that ruins the effect. It also removes the desirable small shifts in frequency that occur as side effect of the wavering time-variation. A more successful solution is to create a 'target delay' within the delay-waver range, toward which each voice travels at each frame through randomized interpolation. Each target delay is chosen randomly for each voice at mutually prime time intervals on the magnitude of a few seconds. At each audio frame, every delay line interpolates toward its target at a random fraction of its current difference, thus preventing any repeating sequence. This staggered 'chasing' effect is unpredictable enough to achieve realistic time and pitch variation without creating unwanted mechanical artifacts. One downside to this process is the high computational demands of calculating random values at the audio rate. Because of this, CPU concerns currently limit the number of duplicate voices to roughly a dozen, however there is much work that could be done to optimize the formula and overcome this limit.

5. PHASE

Phase cancellation is another difficulty that arises from the identical duplication of voices. Although randomized delay shifts detract from the phasing sound characteristic of other Chorus and Flanging effects, it is still a noticeable problem. As a solution, I tried 'cascading' each voice through a series of modified all-pass filters in order to scramble their phases and remove any noticeable pattern. This was somewhat successfully, however under certain parameters there is currently still a noticeable 'robotic' quality to the sound that I believe is due to persistent cancellation. While noticeable, the effect is subtle enough to be ignored with the addition of reverb or when placed within the context of a larger song or mix. In the future, I or others may find a more elegant solution to this problem, however for the time being itâĂŹs a decent fix.

6. SPECTRUM

Ideally, if we were trying to simulate a violin section from a recording of a single violin, for instance, it would be best to pass each duplication through a custom FIR filter that had been designed to replicate the subtle physical acoustic differences found between multiple violins in the real world. Obviously this is impractical for many reasons- not the least of which being computational concerns. The desire is to recreate the almost intangible but distinct complex beating effect that occurs as multiple instruments play a unison line due to the fact that each of them possesses a (subtly) unique frequency response. The differences in the amplitudes of each partial in turn create very complex acoustic interference which we perceive as a queue for the size of an ensemble.

In order to recreate this effect, I found that it was sufficient to assign a single, randomly placed band-reject filter to each duplicated voice. This gives them each a somewhat unique timbre and also promotes these complicated interference patterns when combined with shifts in time and pitch caused by the delay lines.

7. SPATIALIZATION

//figure out how to write up this section

8. RESULTS

//What this piece of software can be used for and how it sounds perceptually. Talk about the different sound examples.

9. FUTURE IMPROVEMENT

Besides increasing efficiency which would allow for many more voices, I'd like to make the system somewhat intelligent. Specifically, I think it should have an option to scale the delay waver based on the general range of the input source. While writing the program I noticed that lower instruments sounded much more "muddy" and dissonant with the same parameters, which leads me to believe that our ears are more forgiving in higher registers. The next version of the program could scale the delay waver accordingly. Additionally, it would be nice to have some sort of delay-waver-waver (meta-waver) that changed over time, either randomly or better yet in response to the music being played. Real instrumentalists are much more/less together depending on the speed, feeling, syncopation, and complexity of a musical passage and it seems feasible that this program could simulate that phenomenon.

10. CONCLUSION

//make this formal and fit it into the narrative

Although there's still much work to be done, currently, it's an effective, simple, and quick way to create large, spatialized audio scenes from very little source material. I think that it's one of the most realistic artificial "ensemble-izer" I've heard to date, and with further improvement it might actually be able to generate what a common listener couldn't distinguish from a real group of performers. The process itself was a practical lesson in how formidable a challenge it is to perceptually decorrelate duplicated audio signals. It was also a psycho-acoustic example of how sensitive our ears are to the subtle nuances of sound- a listener is often able to notice something odd even if he/she couldn't consciously put into words what that peculiarity is.

11. ACKNOWLEDGMENTS

12. ADDITIONAL AUTHORS

13. REFERENCES

[1] M. Dolson. A tracking phase vocoder and its use in the analysis of ensemble sounds. 1983.

[2] D. Kahlin and S. Ternstrom. The chorus effect revisited-experiments in frequency-domain analysis and simulation of ensemble sounds. In *EUROMICRO Conference, 1999. Proceedings. 25th*, volume 2, pages 75–80. IEEE, 1999.

APPENDIX

A. HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to indicate the start of each Appendix, with alphabetic order designation (i.e. the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure *within* an Appendix, start with **subsection** as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 The Body of the Paper

A.2.1 Type Changes and Special Characters

A.2.2 Math Equations

Inline (In-text) Equations.

Display Equations.

A.2.3 Citations

A.2.4 Tables

A.2.5 Figures

A.2.6 Theorem-like Constructs

A Caveat for the TEX Expert

A.3 Conclusions

A.4 Acknowledgments

A.5 References

Generated by bibtex from your .bib file. Run latex, then bibtex, then latex twice (to resolve references) to create the .bbl file. Insert that .bbl file into the .tex source file and comment out the command `\thebibliography`.

B. MORE HELP FOR THE HARDY

The sig-alternate.cls file itself is chock-full of succinct and helpful comments. If you consider yourself a moderately experienced to expert user of LATEX, you may find reading it useful but please remember not to change it.