Compensation of undesired Doppler artifacts in virtual microphone simulations

Nils Peters[1], Jonas Braasch[2]

[1] *Center for New Music and Audio Technologies (CNMAT), UC Berkeley, US, Email: nils@icsi.berkeley.edu*

[2] *Rensselear Polytechnic Institute, School of Architecture, Troy, US, Email: braasj@rpi.edu*

Abstract

Virtual Microphone Control (ViMiC) is a real-time multichannel spatial sound rendering technique based on sound recording principles. In an auditory virtual environment, ViMiC simulates multichannel microphone techniques, resulting in the characteristic Inter-Channel Time Differences (ICTD) and Inter-Channel Level Differences (ICLD) to create the spatial image of a sound scene. When virtual sound sources or virtual microphones are moved, those inter-channel differences are updated in real-time using an interpolation algorithm, which results in a natural Doppler-like pitch shift. However, for musical applications, a Doppler effect is often undesirable. We evaluate the algorithm that aims to avoid such Doppler effects when updating the inter-channel differences. Using various sinusoidal signals, different parameter are evaluated with regards to the spectral distortion.

The ViMiC System

ViMiC [1] is a tool for real-time multi-channel spatial sound synthesis, particularly for concert situations and site-specific immersive installations, and especially for larger or non-centralized audiences. ViMiC is based on sound recording and reproduction principles. A sound recording scenario consists of three main components: 1) sound sources, 2) a recording room, and 3) microphones. The ViMiC algorithm simulates these three components in an auditory virtual environment. Sound sources are defined through their location, radiation pattern, orientation and sound pressure level. Similarly, virtual microphones are characterized through microphone directivity pattern, recording sensitivity, and their position and orientation in the recording room. This virtual recording room causes early reflections and reverberation due to its surface properties and room geometry and size. All parameters can be changed in real-time.

Within the virtual recording room, sound sources and microphones can be placed and moved in 3D as desired (see Fig 1). The propagation path between a sound source and each microphone is simulated accordingly in terms of time-of-arrival (due to distance and the speed of sound), and attenuation, due to a configurable source- and microphone directivity as well as the distance between source and microphone. To increase envelopment and the illusion of a virtual space, ViMiC generates early reflections using a shoe-box room model and the well-known image method. Each image source is rendered according to the time-of-arrival, the distance attenuation, microphone characteristic and source directivity, as described before.

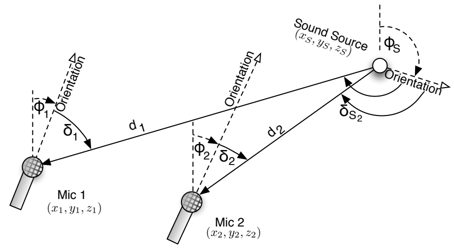

Figure 1: Source - microphone relation in ViMiC.

The Doppler Effect

A continuous change of the source distance engenders the Doppler effect (Eq. 1). In ViMiC, the propagation paths to all microphones (direct sound and all early reflections) needs to be recalculated to update the time-of-flight delays and gain attenuation. A 4-point delay line interpolation is used to continuously adapt the time-of-flight delays for each propagation path separately. This method causes a Doppler-like pitch shift due to a warping of the time axis [2]. Although a natural impression of a moving sound source is created, this effect is often undesired, especially for sustained musical sounds.

$$f = f_0 \cdot \frac{341\,\mathrm{m/s}}{341\,\mathrm{m/s} + v_{\mathrm{source}}} \tag{1}$$

Suppressing the Doppler

Our strategy is based on segmental crossfades (see Fig. 2): the delays of the rendered sound paths remain static until one of the propagation paths has been changed by more than an adjustable threshold. When this threshold is exceeded, the new sound position is rendered (1) and cross-faded (2) with the older position. Finally the algorithm waits for a new threshold (3) and the procedure is repeated (4).

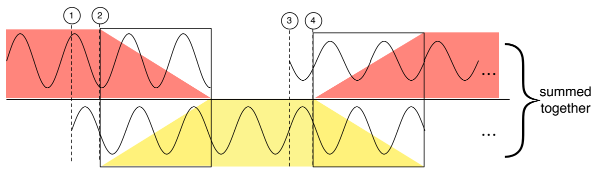

Figure 2: Graphical representation of the algorithm.

Evaluation

Using sinusoidal test sounds, the algorithm is evaluated by comparing the THD+N (total harmonic distortion plus noise) of the Doppler signal and the Doppler suppressed signal. The aim is to improve the THD+N compared to the the Doppler shifted version.

A few adjustable parameters determine this algorithm:

1. Time threshold before a new position is calculated
2. Crossfade shape
3. Duration of the crossfade

Seven crossfade methods (Fig. 3) were evaluated in ten different durations: 2.9, 4.3, 7.2, 11,6 18.9, 29.0, 46.4, 72.6, 114.7 and 182.9 ms. The time threshold was kept at 10 Samples (0.22 ms @ 44.1 kHz). Five sinusoidal sounds (199, 367, 739, 1427, and 3041 Hz) were used. These prime number frequencies are harmonically independent to the crossfade durations. ViMiC was set up to simulate a stereophonic AB microphone configuration with a distance of 1 m between the virtual capsules. No early reflections were rendered. The sound moved along the x-axis on a linear trajectory 6 m in front of the virtual microphones. The speed of the trajectory was 12 m/s (averaged) with a continuous acceleration until the midpoint and a deceleration afterwards until the end.

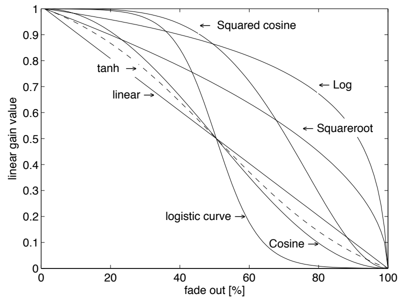

Figure 3: Fade shape types used for crossfading.

Results

Figure 4 shows the spectrogram of one crossfade type in comparison to the untreated Doppler. Here this effect is relatively well suppressed. However, temporal ripples appear in the spectrum which could be associated with roughness. Their form depends on the signal frequency. Figure 5 shows the improvement of the THD+N due to the crossfade methods compared to the rendered Doppler. The *tanh* and *cos* crossfades improved the spectrum most significantly by about 20 dB on average across all tested sounds for longer crossfades. However, this leads to an audible quantization of the trajectory and therefore a loss in spatial accuracy. Also transient sound components (e.g., attacks) can be smeared during the crossfade length. Therefore, shorter crossfades are desired; i.e., for a crossfade length of 29 ms, a *squareroot* crossfade is a better choice. The crossfade algorithm does not perform well for lower frequencies and very short crossfades (lower plot of Fig. 5). Because humans are least sensitive to pitch shifts in lower frequencies [3], a Doppler may not need to be compensated in lower bands depending on the source velocity. Future work focuses on a frequency and source velocity dependent algorithm.

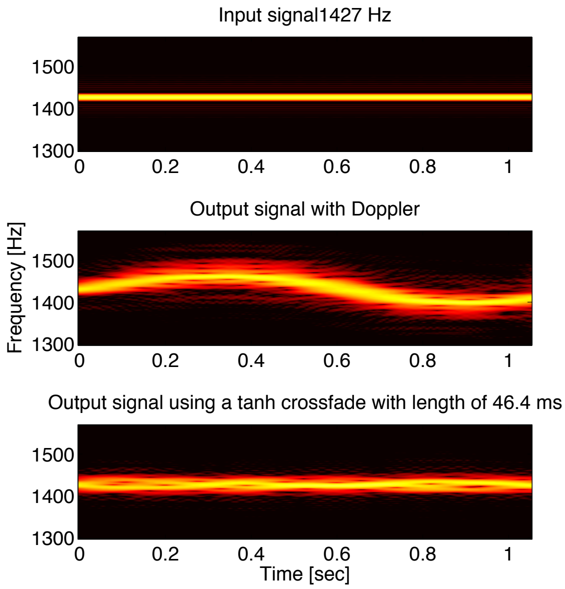

Figure 4: Upper plot: source signal (1427 Hz), middle: rendered microphone signal where delay-line interpolation causes Doppler-like effects. bottom: rendered microphone signal with segmental crossfade using a 46.4 ms *tanh* crossfade.

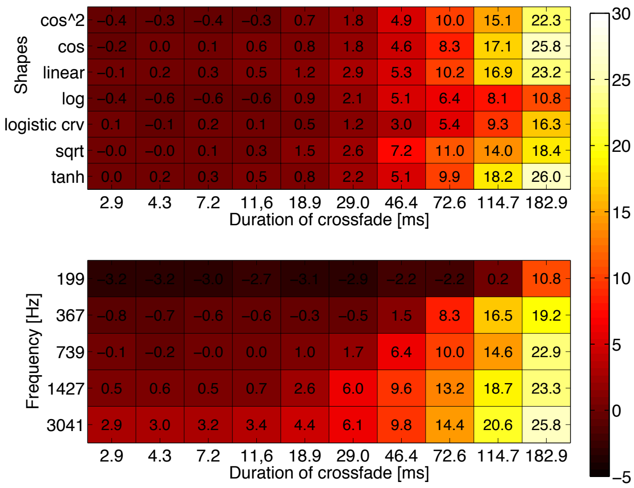

Figure 5: Results of the THD+N measurements with respect to the crossfade duration. Upper plot: Improvement as a function of the crossfade shape (average over all sounds). Lower plot: Improvement as a function of the signal frequency (average over all crossfade shapes).

Acknowledgement

Thanks to Tristan Matthews and Jens Ahrens.

References

[1] J. Braasch, N. Peters, and D. L. Valente, "A loudspeaker-based projection technique for spatial music applications using virtual microphone control," *Computer Music Journal*, vol. 32, no. 3, pp. 55–71, 2008.

[2] J. Ahrens and S. Spors, "Reproduction of moving virtual sound sources with special attention to the doppler effect," in *Proc. of the 124th AES Convention, Preprint 7363*, 2008.

[3] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*. New York, US: Springer Verlag, 3rd ed., 2007.