Real-time auralization of propagation paths with reflection, diffraction and the Doppler shift

Jonas Stienen, Michael Vorländer

Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstr. 5, D-52074 Aachen, {jst,mvo}@akustik.rwth-aachen.de

Abstract

The auralization of early reflections is often implemented using digital signal processing and steady-state filtering by assembling multiple propagation paths into a single filter. To render the Doppler shift in interactive auralization systems, the constraint of linear and time-invariant filtering must be dropped. Instead, resampling during real-time audio processing must be performed. Especially in urban outdoor scenarios, it is not only necessary to account for time variant movement of fast sound sources, but also for dynamics of the built environment. Consequently, if specular reflections and diffraction paths should be considered for real-time auralization, a combination of resampling and filtering must be deployed. A single-input multiple-output variable delay line (SIMO VDL) is suggested to effectively create the Doppler shift for early reflections and to prevent comb filter effects during filter exchange. It consists of multiple interpolating read cursors and subsequent short filtering modules that apply individual spectral attenuation effects caused by medium propagation and boundary interaction.

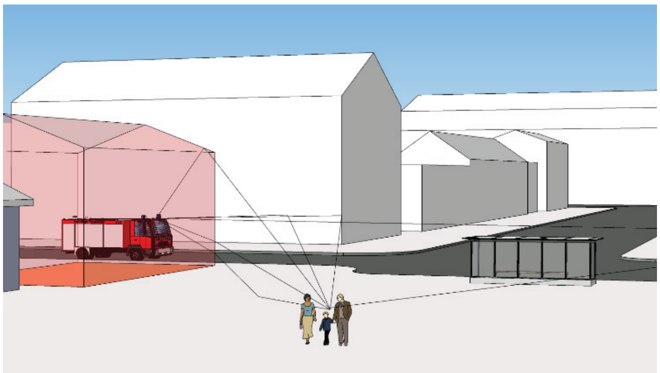

Figure 1: Simplified example of a virtual outdoor environment with buildings, a fast-moving sound source (fire brigade) and a sound receiver (humans). Direct sound is occluded by the red building, hence reflected and diffracted paths dominate the acoustic transmission and individual Doppler shifts per indicated path must be considered.

Sound propagation simulation

To auralize virtual acoustic environments (VAEs), algorithms and models can be employed that approximate the propagation effects of traveling sound waves from a source to a receiver along the wave front normal by means of Geometrical Acoustics (GA) [11], as shown in Figure 1. Especially in the context of interactive indoor and outdoor auralization applications, those GA methods can deterministically find propagation paths that include low-order specular reflections off walls [1] and, with some computational efforts, include diffracted energy as well [2, 9]. The resulting intermediate representation from path finding algorithms can be transformed into time-discrete impulse responses (or transfer functions in frequency domain) that include all steady-state effects such as directivity, medium attenuation, spreading loss, absorption, etc. This approach, however, cannot handle time-variant influences like source and listener movement. The frequency shifts that are hereby caused require resampling [10] using variable delay lines (VDLs) [3, 15]. Adapting to scene dynamics by exchanging filter coefficients or impulse responses and cross-fade the audio streams cause comb filter effects, which become severe when differences in distance increase [12]. For room acoustics, this problem is mostly undetectable [6] and can be suppressed by high filter exchange rates of early reflections and slow sound source movement [7].

Steady-state real-time auralization

Applications that adapt dynamically for changes of the VAE (like moving sources and receivers) operate in real-time, if those adaptions are performed smooth and fast. An impression of immersion is achieved, when human perceptual thresholds are undergone, and the system can readily be used for Virtual Reality applications [11].

This low-latency constraint is demanding and requires fast propagation simulations algorithms, efficient filter coefficient assembly (e.g. FIR filter design) and rapid artifact-free exchange of digital signal processing (DSP) module parameters. Hybrid simulation approaches that handle direct sound, early reflections and diffuse reverberation separately have proven to work sufficiently well when combined with DSP units that implement low-latency block convolvers [6, 12]. The steady-state propagation approach treats sources, receivers and the built environment as entirely static snapshots and, consequently, scene adaptions can only be performed by jumps from one static representation to the next.

Time-variant real-time auralization

For fast moving objects, switching filters that combine all transmission paths can be interpreted as jumps between scenes and result in audible artifacts. Additionally, they lack the possibility to integrate Doppler shifts. Instead, delay lines are commonly used to simulate the medium propagation delay [4, 8, 10]. Interpolating write cursors can effectively auralize the occurring Doppler shift when the source moves relatively against the medium. Furthermore, interpolating read cursors auralize the Doppler shift when the receiver moves relatively against the medium. The Doppler effect becomes perceivable in real-time, if the positional trajectory is sampled at high temporal resolution (e.g. by upsampling and smoothing along spatial polygon courses to match the audio block rate) [10].

This concept of a VDL with interpolating input and output, however, is not able to assemble different propagation paths from a source to a receiver to save DSP load, e.g. by using one convolution unit for all combined paths. The mathematical operators valid for linear time-invariant (LTI) systems cannot be applied to assemble a single (long) impulse response here, because the time-variant write/read interpolation on the VDL represents each propagation path *individually* [8]. The source-related directivity must be applied *before* the audio signal is fed to the VDL (input), while other effects like medium attenuation, diffractions, reflections and receiver directivity must be applied *after* read-out of the VDL (output) and *before* resampling (partly described in [4] and [8]). Wind speed can be integrated easily if the medium is homogeneous and is shifted at a constant velocity.

Anti-causality problem of propagation delay calculation

Calculating the medium propagation delay to determine resampling factors for VDLs with interpolating write cursors as proposed in [8] is only applicable, if positions and orientations of the source are known a-priori or at least ahead in time for a period of the propagation delay $\tau(t)$. The reason is that such a VDL can only represent one wave front direction emitted by a source and thus only applies the Doppler shift for this relative angle against the medium. The angle, however, depends on the future position of the receiver and poses an anti-causality issue: a delay must be introduced to gain causality. Unfortunately, this requirement contradicts with real-time systems as the propagation delay is usually magnitudes higher than the auditory thresholds (in contrast to optics, see [13]).

Establishing causality for Doppler simulation

Auralizing the Doppler shift physically consistent and in real-time is possible if, on the one hand, the variable delay line waives the use of interpolating write cursors and if, on the other hand, the scene representation provides a time history of positions and orientations as proposed in [14] and [13]. This way, quantities like the wave front normal can be found based on the past sound source (ss) position (retarded by τ) and the current sound receiver (sr) position:

$$\vec{r}_N(t) = \frac{\vec{p}_{sr}(t) - \vec{p}_{ss}(t-\tau)}{\|\vec{p}_{sr}(t) - \vec{p}_{ss}(t-\tau)\|_2}$$

The source-to-medium frequency scaling can be formulated based on velocities of sound source and medium

$$\frac{f_m}{f_{ss}} = \frac{1}{1 - \vec{r}_N(t) \cdot \frac{\vec{v}_{ss}(t-\tau) - \vec{v}_m}{c}}$$

as well as medium and sound receiver

$$\frac{f_{sr}}{f_m} = 1 - \vec{r}_N(t) \cdot \frac{\vec{v}_{sr}(t) - \vec{v}_m}{c}$$

with c the speed of sound and the dot product being the vector projection. In the suggested causal design, the Doppler shift of both the source-to-medium transition *and* the medium-to-receiver transition must be combined to determine the resampling factor:

$$r = \frac{f_{sr}}{f_{ss}} = \frac{1 - \vec{r}_N(t) \cdot \frac{\vec{v}_{sr}(t) - \vec{v}_m}{c}}{1 - \vec{r}_N(t) \cdot \frac{\vec{v}_{ss}(t-\tau) - \vec{v}_m}{c}}$$

Resampling by r must be applied during interpolating read-out.

The proposed modification violates the above stated processing chain in two ways: firstly, the source directivity cannot be applied *before* interpolating write and secondly, the propagation filters along the path (reflections, diffractions, air attenuation) are not modifying the *scaled* frequencies that are physically present in the medium until the signal is picked up by the interpolating read cursor at the receiver position. Fortunately, both problems can be circumvented by considering the Doppler shift at both medium transitions for subsequent units: the directivity and the path filter coefficients must be reversely scaled to consistently match the attenuation of frequency bins if the processing would be performed in the valid order.

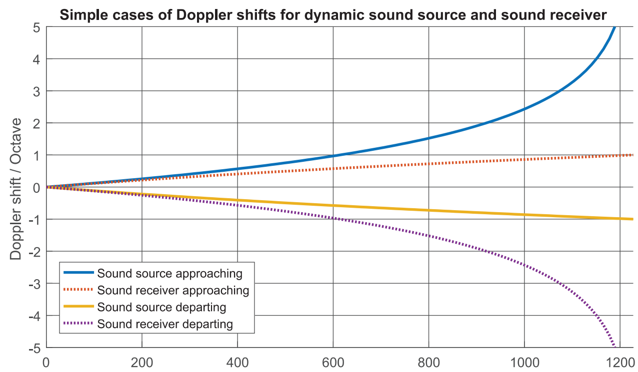

Figure 2: Simplified Doppler shifts per Octave for approaching and departing source and receiver.

Consider the following simple example with static source, static receiver and constant downwind from receiver point of view: source signals are shifted towards higher frequencies at source-medium transition in the direction of the receiver. Signals are then shifted back towards lower frequencies at medium-source transition (compare Figure 2). The combined frequency scaling is reciprocal, and the resampling factor will be 1 resulting in no audible Doppler shift. However, the air attenuation filter will have a stronger low-pass character. To account for this behavior in the proposed signal processing, coefficients are pushed towards lower frequencies. Reason is that the physical sound pressure wave in the medium contains higher frequencies upwind due to the Doppler squeezing, which in turn leads to stronger attenuation along the path (and additionally to a longer travel time).

Limits of Doppler simulation by resampling with time-discrete audio processing

Making Doppler shifts audible by *squeezing* time-discrete audio signals results in aliasing. Frequencies near the Nyquist limit are pushed above maximum representable frequency when resampling with $r > 1$. The input signal therefore must be low-pass filtered to suppress artifacts [5]. Because the design of the anti-aliasing filter depends on the resampling factor, a direct link between scaling and source/receiver movement can be derived. The anti-aliasing can be implemented by an adaptive filter prior to resampling, which requires one additional processing unit for each path. With some preliminary considerations with respect to source and receiver velocities and maximum perceivable frequency, the additional unit can be saved without degradation. The input signal can be low-pass filtered *before* feeding the VDL or can simply be band-limited by design requirement in order to save the anti-aliasing filter at once. For this approach, figure 3 shows shifted Nyquist frequency over approaching sound source speed. The curves indicate aliasing limits for different sampling rates. For instance, at $44.1kHz$ and with $18kHz$ being the cut-off frequency of low-passed input signals, the maximum allowed velocity is $200\ km/h$, which appears sufficient for room acoustics and city acoustics, but not for high speed trains and air-crafts. The situation can be relaxed further if either higher sampling rates are used, or the cut-off frequency is lowered.

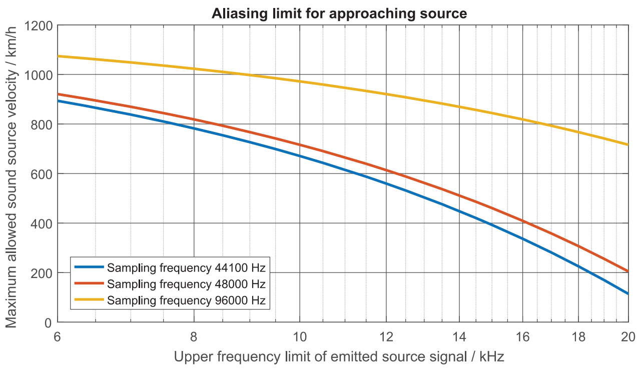

Figure 3: Aliasing limit for approaching source for different sampling rates. For band-limited signals, the maximum allowed speed can be depicted on the curves.

In turn, *stretching* time-discrete signals potentially shifts non-representable frequency content into the representable range below Nyquist limit. This case is most likely to occur when a fast-moving receiver departs from the sound source, in which case air attenuation would almost instantly dampen extremely high frequency content and render them imperceptible. Hence, this theoretical side note can be considered a topic without real acoustic relevance.

Single-input multiple-output variable delay line

A DSP design is proposed that uses a single SIMO VDL with non-interpolating source input and dedicated receiver output processing for each path, which is finally superposed for the receiver. The main concept is based on the idea that all relevant acoustic features of transmission (directivity, reflections, diffractions, air attenuation, spreading loss and the Doppler shift) can be applied at latest time possible, namely at time t when the receiver is picking up the emitted sound waves from the transmitting medium. The delay line is merely a buffering unit storing and transporting the emitted source signal at sound speed c to the receiver for all geometric propagation paths.

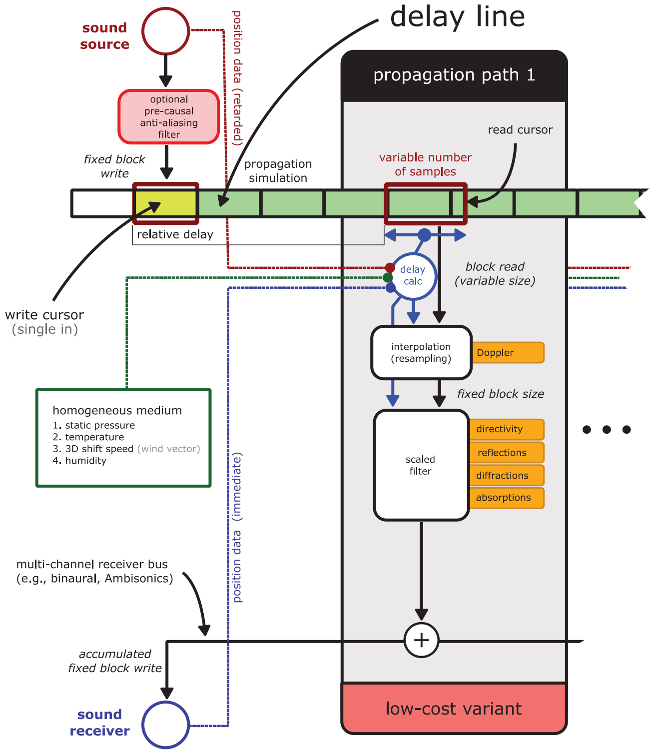

Figure 4: SIMO VDL for a band-limited input, combined Doppler resampling and frequency scaled filtering (only one read cursor DSP block depicted).

The solution is based on the considerations stated above and uses a single non-interpolating write cursor (see Figure 4). With interpolating read cursors (multiple output), reflected or diffracted propagations paths are simulated that apply individual Doppler shifts determined by resampling factors against image source positions [1], which may have opposite scaling directions for the same source.

If the sound source signal is band-limited per definition (or low-pass filtered beforehand) and movements are not violating the aliasing limit (see Figure 3), the SIMO VDL and subsequent filtering can be applied very efficiently. The interpolating read cursor stretches or squeezes the time signal, and the receiver filter component is parametrized by adaptively scaled coefficients for directivity and further propagation filters. The block diagram of this low-cost variant is depicted in Figure 4 and is a solid choice for room acoustics and city acoustics without fast driving vehicles. Additionally, the implementation is straight forward.

Another variant that requires higher processing performance uses a multi rate approach (see Figure 5, left). Aliasing issues are reduced by upsampling the VDL read-out with a given multiplicator (e.g. doubled). The allowed speed constraint for source and receiver movements is accordingly lowered (cmp. allowed speeds at $48kHz$ sampling against doubled

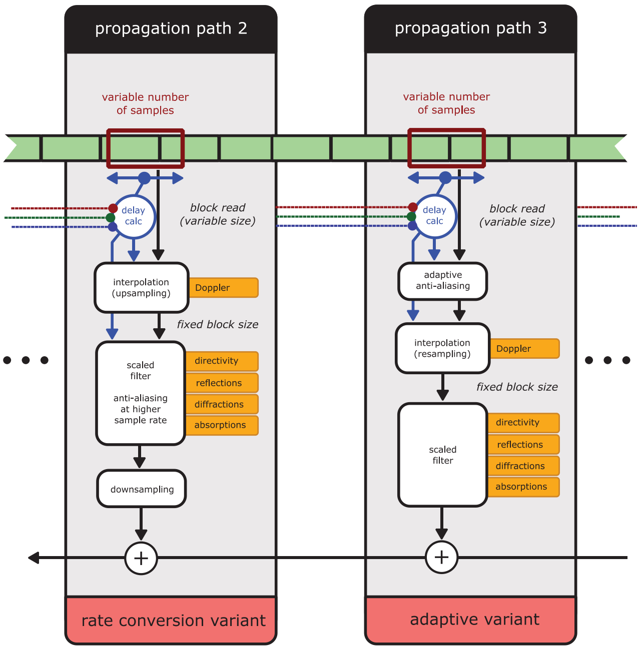

Figure 5: SIMO VDL read-out block diagrams for more relaxed aliasing issues using upsampling (left) and adaptive anti-aliasing filter (right).

sampling at $96kHz$ in Figure 3). However, the higher sampling mode and the inevitable low-pass filter for downsampling [5] increase required processing power.

The physically most reasonable DSP design includes an adaptive low-pass filter unit to avoid aliasing based on actual source and receiver movements. This design concept is based on an additional filter unit that requires variable block length operation and is inserted between VDL non-interpolating read-out and the interpolation routine (see Figure 5, right).

Both the upsampling and the adaptive variant are valid choices to handle fast moving sound sources like high speed trains and aircrafts, however the adaptive filtering consequently avoids aliasing while (at least a fixed) sampling rate conversion presupposes speed limits.

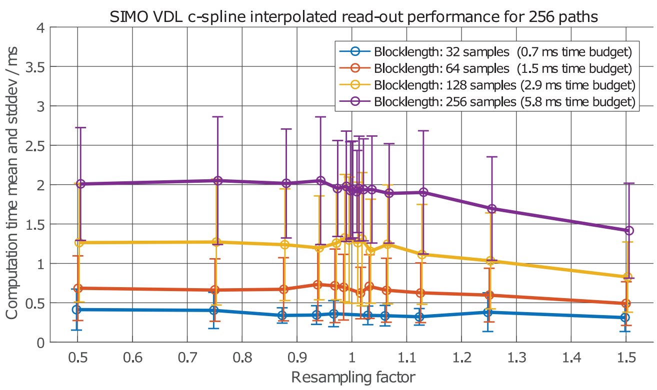

Figure 6: SIMO VDL cubic-spline interpolated read-out benchmark for 256 propagation paths and N = 10,000 repetitions on Intel Core i7-4770 CPU@3.40Ghz 64-bit Microsoft Windows 10 using a binary compiled with Microsoft Visual C++ Compiler Version 2012.

Performance and conclusion

The performance of the interpolating read cursors have been benchmarked (without any filtering). For 256 paths, cubic-spline read-outs at different resampling factors and block lengths have been executed and the result is depicted in Figure 6. Half of the real-time processing time budget is required by the SIMO VDL in this case. Therefore, subsequent filtering units for all other propagation effects despite Doppler shifts will impose the limiting factor for real-time operation, not the SIMO VDL processing.

References

[1] Allen, J.B. and Berkley, D.A. 1979. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* 65, 4 (1979), 943–950.

[2] Calamia, P.T. et al. 2005. Integration of edge-diffraction calculations and geometrical-acoustics modeling. *Proceedings of forum acusticum* (2005).

[3] Hacihabiboglu, H. 2003. A fixed-cost variable-length auralization filter model utilizing the precedence effect. *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.* (2003), 1–4.

[4] Lokki, T. and others 2002. *Physically-based auralization: design, implementation, and evaluation*. Helsinki University of Technology.

[5] Ohm, J.-R. and Lüke, H.D. 2007. *Signalübertragung - Grundlagen der digitalen und analogen Nachrichtenübertragungssysteme*. Springer.

[6] Pelzer, S. 2008. *Interactive real-time auralization of complex virtual environments*. RWTH Aachen.

[7] Schröder, D. 2011. *Physically based real-time auralization of interactive virtual environments*. RWTH Aachen.

[8] Smith, J. et al. 2002. Doppler simulation and the Leslie. *Proceeding of the Workshop on Digital Audio Effects, DAFx-02, Hamburg, Germany* (2002), 188–191.

[9] Stienen, J. and Vorländer, M. 2017. Geometry-based diffraction auralization for real-time applications in environmental noise. *The Journal of the Acoustical Society of America*. 141, 5 (2017), 3536–3536.

[10] Strauss, H. 1998. Implementing Doppler Shifts for Virtual Auditory Environments. *Audio Engineering Society Convention 104, Amsterdam, NL* (May. 1998).

[11] Vorländer, M. 2011. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer.

[12] Wefers, F. 2015. *Partitioned convolution algorithms for real-time auralization*. Logos-Verl.

[13] Wefers, F. and Vorländer, M. 2018. Flexible data structures for dynamic virtual auditory scenes. *Virtual Reality*. (2018), 1–15.

[14] Wefers, F. and Vorländer, M. 2014. Zeitvariante Beschreibung virtueller Szenen für die Echtzeit-Auralisierung instationärer Schallfelder. *Jahrestagung Deutsche Gesellschaft für Akustik (DAGA 2014), Oldenburg, Germany* (2014).

[15] Zölzer, U. 2008. *Digital audio signal processing*. John Wiley & Sons.