SIZE MATTERS: NEW WAYS OF WORKING WITH SCALE MODELS

BFG Katz   Institut d'Alembert, Sorbonne University, CNRS
P Stitt   Institut d'Alembert, Sorbonne University

1   INTRODUCTION

This project aims to investigate the use of new scale model tools in room acoustics research. While acoustic studies using scale models have been part of the toolbox of acousticians for quite some time, back to the use of ripple tank over a century ago, recent advances in transducer and data acquisition hardware as well as computational power have opened up the possibility for high fidelity scale model measurements as well as real-time audio processing. This study presents some recent advances in scale model acoustic research developed by our team. Microphone miniaturisation, MRI scanning, and high-resolution 3D printing have been used to create a scaled version of a Neumann KU-100 dummy head for binaural acquisition. Real-time signal processing has been leveraged to develop a prototype audio processor plugin which is capable of transforming full-scale audio streams to scaled-audio and transforming the captured scaled-audio back to full-scale for a real-time listening experience within the scale model. The combination of these developments moves scale models from being a limited omni-directional off-line and somewhat laborious analysis tool into a more realistic and interactive design experience platform. Ultimately, the goal is to advance the field of room acoustics by exploring the potential of these new tools to improve acoustical testing, design, and analysis.

In 1934, F. Spandöck introduced subjective testing of acoustics of rooms based on 1:5 scale model auralisations.[1] Changing the playback rate of a magnetic tape recorder, speed-up audio samples such as music or speech were reproduced using a loudspeaker inside the scale model, providing the frequency transposition. This article presents characterisation of hardware (sources and receivers) for use in scale model auralisation. It then proceeds to present a real-time audio processing plugin which provides up/down-scaling of audio for listening dynamically within a scale model. Processing steps to account for the frequency responses of different transducers is included to improve the audio quality of the result. Significant effort has been made to optimize latency for real-time use.

2   HARDWARE

Rendering a scale model auralisation that covers the whole audible bandwidth requires equipment that can record and playback at frequencies at k-times the upper limit of the human auditory system. With digital equipment the sampling frequency f_s of the audio interface provides a hard limit to the maximum reproducible frequency. The chosen sampling frequency allows for a maximum reproduction frequency of $f_s/2k$. It should be noted that the anti-aliasing filters used by pro-audio (versus laboratory equipment) interface manufacturers could have a non-negligible impact on the spectrum of the signal, since there may be a gradual roll-off above the audible frequency range. Many audio interfaces, such as the RME BabyFace Pro[i], support sampling frequencies of up to 192 kHz. For a scale factor of 1:10 this allows for a maximum auralisation frequency of 9.6 kHz.

Since the frequency range between the upper limit of the auditory system and the maximum recording frequency is usually of little interest for traditional audio applications, measurements are made to characterise and correct the frequency response of sources and receivers across the whole usable frequency range. Impulse responses were measured using the swept sine technique[2] for frequencies spanning 1 kHz to 90 kHz. An RME Babyface Pro audio interface was used with a sampling frequency of 192 kHz. Since the accuracy of the ultrasonic frequency range is of critical interest, a loopback recording was made of the audio interface. The frequency response of the loopback performance

---

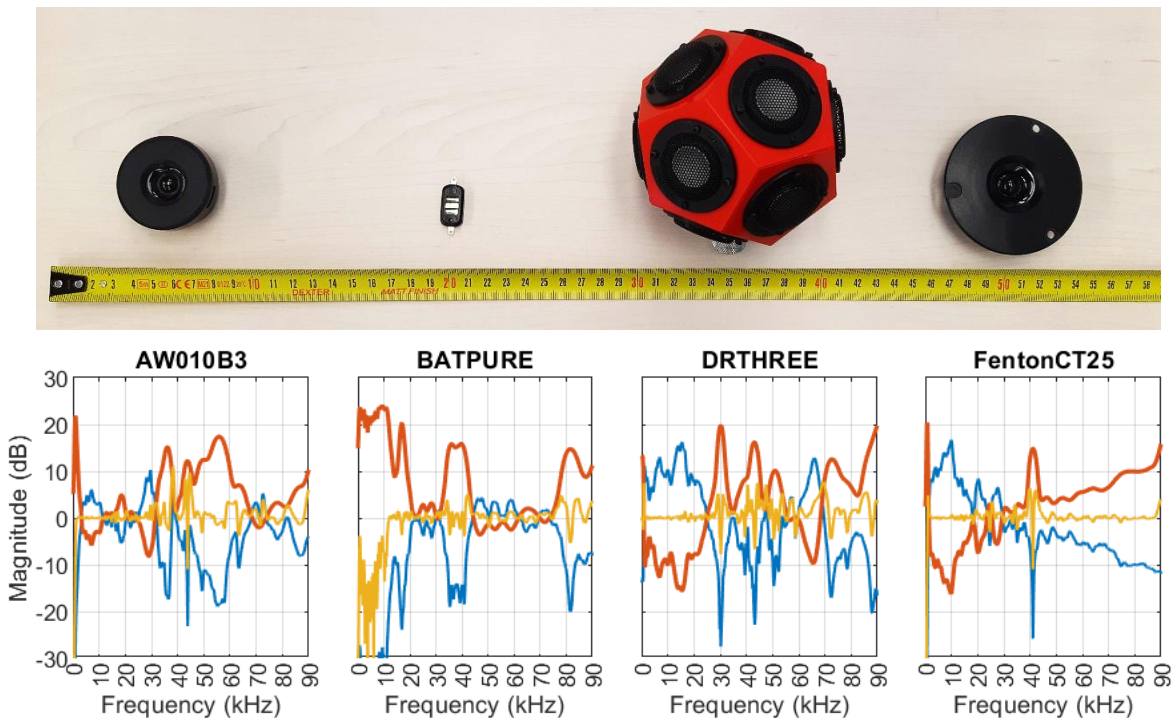[i] RME Babyface Pro FS. https://rme-audio.com/babyface-pro-fs.html

Figure 1 – (Top) The 4 measured tweeters with a measuring tape for scale. (Bottom) their magnitude frequency responses: measured (**blue**), inverse with semitone smoothing (**red**), EQ'd, *i.e.*, original impulse response convolved with the inverse filter (**yellow**).

showed a slight magnitude roll-off with increasing frequency, so the loopback impulse has been deconvolved from all measurements.

2.1 Source Measurements

Four tweeters were measured to evaluate their performance for use with scale model auralisation. Measurements were made with a 1/8" GRAS reference microphone, specified as having of a flat frequency response over the measured frequency range.

The four sources that were measured were: Audax AW 010 B3[ii] (dome tweeter, 10 mm diameter), Batpure "super tweeter" (PVDF ultrasonic tweeter)[iii], Dr-Three dodecahedron speaker[iv], and Fenton CT-25 (dome tweeter, 12.5 mm diameter). Except for the Dr-Three, all were unmounted without any housing or added baffle. It is noted this is not the use intended by the manufacturers. The intended use-case in the scales models is without mounting or baffles, to keep their size small, which may affect their frequency response in the audible range.

The magnitude responses of the four sources are shown in Figure 1, normalised to 0 dB mean level between 20 kHz and 60 kHz. As expected, the frequency responses are far from being flat. The AW010B3 has significant variations across the frequency range but without any pattern of roll-off with increasing frequency. The Batpure has little signal below 15 kHz and a wide notch between 30 and 40 kHz, being otherwise relatively flat up to 80 kHz. The large Dr-Three omnidirectional source has strong peaks and notches across the frequency range. The Fenton CT25 exhibits a broadly constant roll-off as a function of frequency, along with a notch just above 40 kHz.

To correct for the frequency response of the sources, inverse filters were computed as minimum phase FIR filters: (1) inverse spectrum calculated with Tikhonov regularisation, (2) inverse spectrum magnitude smoothed using semitone smoothing, then (3) resulting zero-phase spectrum converted

---

[ii] https://www.lamaisonduhautparleur.com/en/speakers-discount/1709-aw-010-b3-4ohm.html

[iii] TAKET-BATPURE Super Tweeter. http://www.taket.jp/batpure/batpure.html.

[iv] 3D Sound Speaker 3D-032. http://www.dr-three.com/products/m3d032.html.
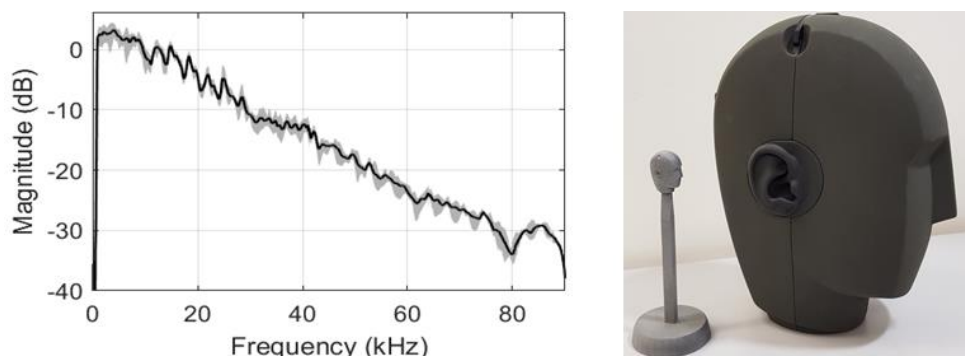
Figure 2 – (left) The mean magnitude response of 4 miniature Feichter M1 microphones (for use in scaled dummy-head). The mean response is bordered by the minimum and maximum magnitude response across all measurements. (right) Photo of KU-100, scaled model and actual size.

to minimum-phase and transformed back to the time-domain. Figure 1shows the magnitude response of each inverse filter, as well as result of the measured impulse responses convolved with the inverse filter. The EQ'd result is a significantly flatter and smoother across the spectrum. The regularisation and smoothing approach was used to avoid creating filters having strong peaks attempting to remove narrow notches. As such, the resulting corrected signal is not completely flat but in all cases is an improvement compared to the un-corrected signal.

2.2 Microphone Measurements

Four example miniature microphones (Feichter M1[v], based on Knowles with phantom power preamp) were measured using the Fenton CT25 tweeter, deconvolved with reference CT25 and GRAS measurements. Tikhonov regularisation was used for the calculation of the inverse frequency response during the deconvolution. As such, assuming a flat frequency response of the GRAS reference microphone, the impact of the tweeter is removed, leaving only that of the microphones. Figure 2 shows the mean response across all four microphones, as well as the minimum and maximum values around the mean. The microphones are relatively consistent between microphones. They are in the range of repeated measurements when removing and replacing the microphones from the stand. The measurements show a consistent roll-off with increasing frequency until they finish with almost 40 dB of attenuation at the maximum measured frequency.

Since the frequency responses of both the tweeters and the microphones deviate significantly from flat across the bandwidth it is highly likely that a compensation EQ will need to be applied to improve the overall frequency response. While necessary, it should be noted that compensation of such extremely deviant spectra will also be applying a significant boost to noise at the upper end of the frequency range. For example, compensating the >30 dB attenuation of the Fenton CT25 along with the Feichter M1 microphones' own >30 dB of attenuation at the top of the frequency range will result in more than a 60 dB boost in the background noise at the top of the frequency range. Therefore, care must be made to select the best possible equipment to allow for the highest quality auralisation.

2.3 Scale Model KU-100 Dummy Head

A 1:8.5 scale KU-100 dummy head (Figure 2) was created using high resolution scans to capture the geometry which was then transformed to a 3D mesh and scaled. It was designed such that two Feichter M1 could be placed inside with flexible seals around the interior of the ear canals[vi] and then 3D printed[vii] (Multijet fusion, Nylon PA12, 60µm HD resolution). The scale dummy head measured with an angular resolution of 5° on the horizontal plane using an automated turntable.

---

[v] http://feichter-audio.com/produits/captations/m1/
[vi] Scale Head fabrication project, https://pyrapple.github.io/pages/scale-head.html
[vii] Sculpteo 3D printing service, https://www.sculpteo.com/
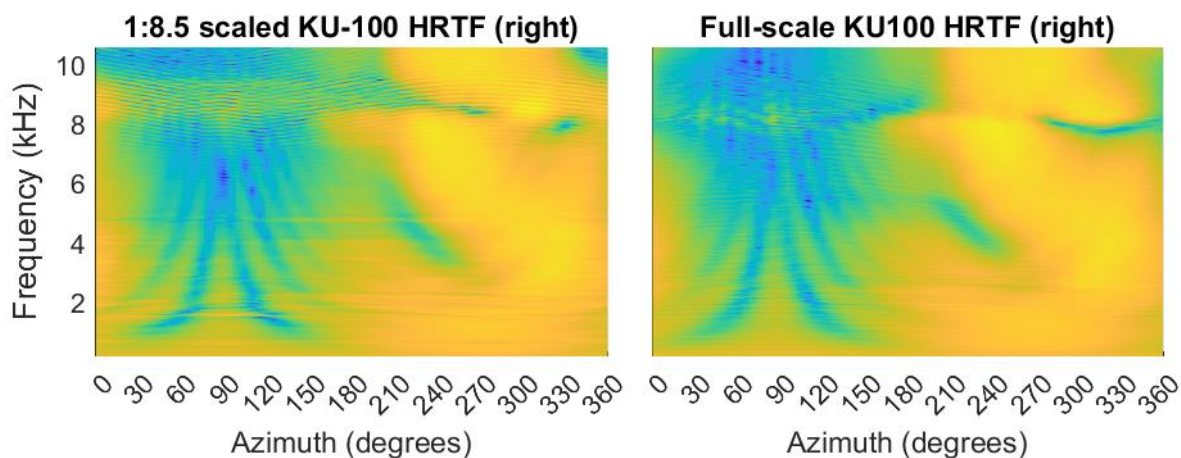
Figure 3 - Comparison of the HRTFs of the scale model KU-100 and the full-scale measurements for the horizontal plane.

Figure 3 shows the diffuse field equalised HRTF (for the right ear) alongside measurements of the full-scale KU-100[3]. There is broadly good agreement in the HRTFs. The scaled dummy head measurements show some ripples in the frequency domain which could be due to the resonances, slight gaps in attaching the two sides of the dummy head, or slight imperfections in the rubber seals at the ears which are planned to be improved in the next version. Above 8 kHz the scale model measurements show some ripples in the frequency axis, likely due to noise resulting from low signal levels due to microphone and tweeter responses, as mentioned in Section 2. Overall, up to 8 kHz the scale model provides a good approximation of the full-scale model. Minimum phase diffuse field FIR filters were calculated to allow for real-time diffuse field compensation.

3 REAL-TIME AUDIO PLUGIN: SMALLROOMZ

3.1 Previous Implementations

Spratt and Abel[4] proposed an implementation for real-time scale model auralisation. They present an overview of the essentials of the method without specifying some of the specifics. In their method, they apply an anti-aliasing filter before down-sampling the signal by selecting every k-th sample. The down-sampled signal is then zero-padded and output to the scale model where it is recorded. The signal returned from the scale model is then reconstructed using the overlap-add method, so that the reverberation tails recorded for each buffer overlap with the direct signal of the next buffer. The signal is up-sampled by inserting $k - 1$ samples between each sample before finally having an anti-imaging filter applied to the output. The anti-imaging filter is the same as the anti-aliasing filter.

This work was expanded upon by Delcourt et al.[5], which was implemented in Max 8[viii]. They formalised the method of Spratt and Abel[4], providing information on the length M of the down-sampled buffer segments based on the scale factor and reverb time of the scale model. They specified that for a given scale model reverberation time RT'_{60} the following inequality must hold:

$$RT'_{60} \leq \frac{M(k-1)}{f_s}.$$

If M is too short then the reverberation tail may still be audible when the next buffer is played, invalidating the assumptions required for the overlap-add processing and leading to audio artefacts. In addition, they included an EQ to allow the user to correct for the often-non-ideal frequency responses of the equipment used in the ultrasonic frequency range. It should be noted that this RT'_{60} time parameter also includes the propagation time from source-to-receiver.

The method performs well when there is no background noise present in the scale model. However, if any background noise is present (such as air conditioning) this leads to clicks whose frequency is

---

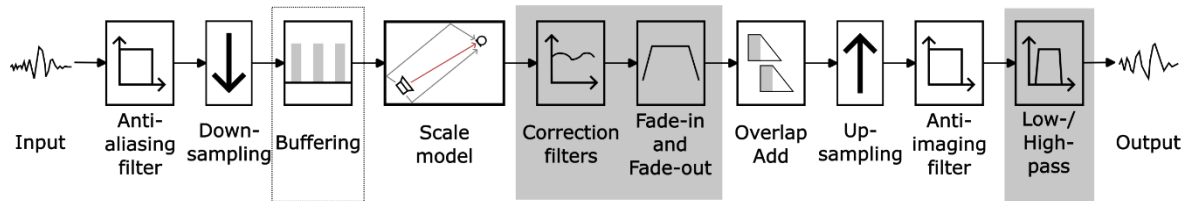[viii] Cycling 74 Max 8. https://cycling74.com/products/max.

Figure 4 - The signal schematic for the real-time scale model auralisation plugin implementation. The new elements are indicated by the grey background. The block with the dotted outline has been modified compared to the Delcourt et al.[5] implementation.

related to the size of the buffers used for the overlap-add buffering. Essentially, the background noise creates a discontinuity in the signal which results in a click at the start and end of each overlapped segment. It was suggested that some windowing could mitigate this, but was left for future work.

Any implementation of this real-time algorithm will have an inherent latency linked to the buffer size, which is itself a function of the scale factor, sampling frequency, and reverberation time of the scale model. The Delcourt et al. implementation would output a down sampled block after receiving kM input samples, *i.e.* after collecting M samples in the output block. This represents the minimum latency achievable before the signal is output to the scale model (excluding the audio IO latency).

3.2   New Plugin Implementation

A new implementation of this real-time scale model auralisation has been implemented as a Steinberg VST3[ix] plugin. It was written in C++17 using the JUCE 7 framework[x], allowing it to be used in any host software capable of playing incoming real-time audio through the plugin in real-time, such as Max. The new implementation also attempts to minimise the latency as much as possible, as well as providing parameters to the user that can be adjusted when in the presence of background noise.

3.2.1   Plugin Additions and Improvements

The plugin uses the same signal flow architecture as Delcourt et al.[5] (tested to 1:5 scale) and Spratt and Abel[4] but adds some additional blocks, as well as modifying the down-sampled buffer output slightly. The signal flow is shown in Figure 4. The new elements are indicated by the grey boxes.

The first improvement over the Delcourt method is in the "Buffering" block, indicated by the dotted box outline in Figure 4. In this implementation the down-sampled audio output begins being output even before all M samples have been buffered. The first sample is output to the scale model such that the last sample of the buffer sent to the output directly after it has arrived. The latency of the down-sampling processing is therefore reduced from kM samples to $(k-1)(M-1)$ samples.

The second addition to the signal flow is the "Correction Filters" block, allowing for the correction filters obtained in Section 2 to be applied, improving the frequency response of the auralisation. The correction filters are implemented as 512 sample minimum phase FIR filters, sampling rate of 192 kHz, having a length of 2.67 ms. Additional correction filters can be imported as .wav files.

The next improvement is the addition of the "Fade-in and Fade-out" signal block before the overlap-add processing, applying a cosine-window to the start and end of the signal that is captured in the scale model. To avoid windowing the start of the desired signal, a delay of the fade-in window (t_{in} sec) is applied before the down-sampled signal is output to the scale model. This ensures that the fade-in is only applied to background noise captured before the desired signal. The fade-out (t_{out} sec) is applied at the end of the buffer. In order to compensate for the time for the fade-in and fade-out, the inequality that M must satisfy becomes:

$$RT'_{60} + t_{in} + t_{out} \leq \frac{M(k-1)}{f_s}.$$

[ix] Steinberg 3rd Party Developer SDKs https://www.steinberg.net/developers/.
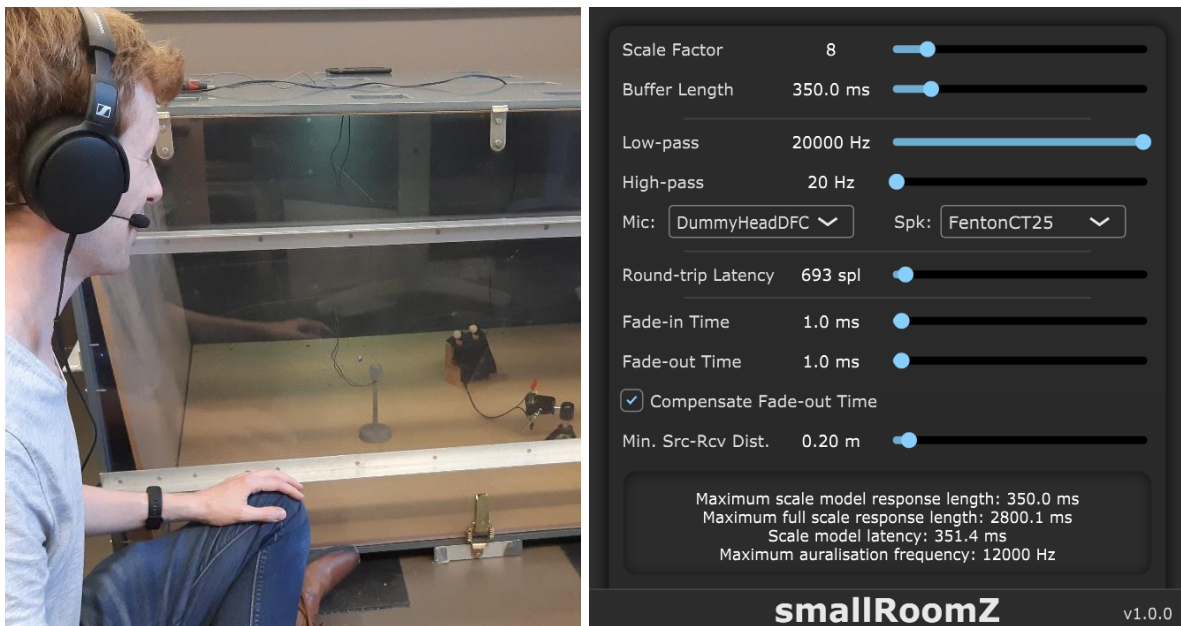[x] JUCE Framework https://juce.com/.

Figure 5 - (left) A user speaking & listening to their voice though the scale KU-100 dummy head in a simple PVC/MDF scale model. (right) smallRoomZ user plugin interface showing all controls for parameters and information based on the current settings at the lower portion of the interface.

The fade-in and -out therefore leads to a double increase in the latency of the system: once from the delay of t_{in} before outputting to the scale model and again by the increase of M to accommodate the length of the windows. However, the time for the fade-in window can potentially be compensated if the minimum source-to-receiver distance d_{in} in the scale model is known. Since there is no signal other than background noise during this propagation time, the fade-in can be applied during this period, if it is long enough. Thus, the inequality for M becomes:

$$RT'_{60} + \max(0, t_{\text{in}} - d_{\text{in}}/c) + t_{\text{out}} \leq \frac{M(k-1)}{f_s},$$

where c is the speed of sound. If $t_{\text{in}} \leq d_{\text{in}}/c$ then there is no latency associated with the fade-in. In this case, the corresponding delay added to the output to the scale model is also reduced to $\max(0, t_{\text{in}} - d_{\text{in}}/c)$ samples.

The final block added to the signal processing chain is simple high- and low-pass IIR filters. These are applied to the final output after the anti-imaging filter, intended to allow the user some final control of the bandwidth of their signal, depending on the equipment used.

3.2.2   Plugin User Parameters

The user interface for the plugin is shown in Figure 5 showing the parameters available:
- **Scale Factor**: The scale factor k of the scale model being used.
- **Buffer Length**: The length of the zero-padding applied to the down sampled audio buffer. This should be long enough to capture the entire reverberation tail of the scale model, along with source-to-receiver propagation delay.
- **Low-/High-Pass**: Low- and high-pass filters applied to the final auralisation to remove unwanted frequency regions.
- **Microphone/Speaker EQ presets**: Selection of the microphone and speaker EQ presets to be applied to the signal returning from the scale model.
- **Round-trip Latency**: The delay for a signal to output and then return introduced by the audio interface. It is primarily governed by the I/O buffer size B specified in the audio driver settings, but can include additional hardware-related latency beyond waiting for the buffers to be filled. It should be measured using a loop-back technique and supplied in samples. Regardless of the user input a value of $2B$ (theoretical round-trip latency excluding hardware-related delay)

is used internally. The round-trip latency is used to initialise the read-heads of the overlap-add processing such that a read-head reads the return from the scale model at the correct time, ensuring minimum latency.

There are also advanced parameters for users who wish to further minimise the latency of the system, such as in cases where a performer might wish to hear themselves in the room, rather than someone listening from an audience position. These parameters also include control of the fade-in and -out windows that can be adjusted in the case of background noise. The advanced parameters are:

- **Fade-In Time**: The length of the fade-in window t_{in} used to suppress transient noise due to the presence of the background noise.
- **Fade-Out Time**: As above, but for the fade-out time t_{out}.
- **Compensate Fade-out Time**: If this parameter is activated then the length of the fade-out is added to the zero padding of the down sampled buffer. This is to ensure that the fade-out is not applied to the desired portion of the reverberation tail specified by Buffer Length, causing a truncation of the reverberation. Disabling this parameter reduces the latency but may introduce audio artefacts due to the premature fading of the reverb tail.
- **Minimum Source-Receiver Distance**: This specifies the closest distance the source will have to the receiver in the scale model d_{in} in metres. If a source-receiver distance greater than zero is specified and the source is brought closer than this distance, the resulting fade-in window will be applied to the direct sound, creating artefacts in the auralisation.

Based on the values set for each of these parameters, along with the sampling frequency, several pieces of information are displayed to the user at the bottom of the interface. The maximum impulse response length to avoid truncation (including source-to-receiver propagation time) in the scale model along with the corresponding scaled time are displayed so the user can ensure they have set the buffer length to capture the desired length of the reverberation tail. The latency between the signal being input and sent to the scale model is also displayed in milliseconds. As a reference, the maximum reproducible frequency of $f_s/2k$ is displayed, informing the user of the signal bandwidth.

3.2.3 Total System Latency

The plugin has been designed to work with a real-time input stream and to play back the auralisation from the scale model also in a continuous real-time output stream. In the case where a performer might want to listen to their own performance live in a room, latency must be minimised. Depending on the scale factor and Buffer Length parameters, it might not always be possible to achieve a latency low enough for this kind of use. However, for the cases where it could be possible, options are available to the user to minimise the latency as much as possible.

Latency is present at many points in the system, both from the signal processing and from the audio interface used for playback within the scale model itself. The first point at which latency is applied is the anti-aliasing filter applied to the input signal. An IIR filter implementation was chosen, as a linear-phase FIR filter would have introduced significant latency due to the steep cut-off.

The second, and most significant source of latency in the DSP, is waiting to full the buffer with enough samples to be input before the signal can be down sampled and output to the scale model. As mentioned in Section 3.2.1, this is $(k-1)(M-1)$ samples.

If a fade-in time >0 s has been specified, then this time will be added as a delay to the down sampled audio before output to the scale model, adding additional latency to the system. The total delay depends on the user-supplied minimum source to receiver distance. The total fade-in latency is $(\max(0, t_{\text{in}} - d_{\text{in}}/c))$ samples.

The non-DSP-related delay is due to the round-trip latency of the audio interface and the propagation time between the source and the receiver in the scale model. The round-trip latency is considered as part of the system latency, since it is related to the playback system, but the source-receiver propagation delay is more properly considered to be an inherent part of the auralisation.

The read-head for the up-sampling before the output is initialised such that it will arrive at the first sample of the return at the same time as the first sample of the return from the scale model, assuming the round-trip latency has been correctly supplied. In the case that the round-trip latency value is set

to too low a value, the read-head may already be in front of the signal and the output will be delayed until the next read-head arrives. As such, it is important that round-trip latency be accurately supplied. The final latency is the anti-imaging filter for the up-sampled audio, same as the anti-aliasing filter.

The total latency of the system, *excluding* the frequency dependent delay of the anti-aliasing and anti-imaging IIR filters, is therefore

$$\text{Total latency} = (k - 1)(M - 1) + \lceil f_s(t_{\text{RTL}} + \max(0, t_{\text{in}} - d_{\text{in}}/c)) \rceil$$

where t_{RTL} is the round-trip latency in seconds. Since the reverberation time of the system, which governs M is a function of the physical scale model, as is k, the first term cannot be reduced for a given model. The latency must be reduced by using the minimum audio interface buffer size possible and, ideally, ensuring that $t_{\text{in}} \leq d_{\text{in}}/c$.

3.2.4   Additional Background Noise Consideration

Due to the overlap-add algorithm used, the signal-to-noise ratio is reduced. As the scale factor increases, and thus the number of overlapping segments, the noise level will also increase. There will be k overlapping buffers at all points in time which means that the noise level will increase by 3 dB for each doubling of the scale factor. This provides further motivation to ensure the auralisation is made in the quietest possible environment, particularly at high scale factors.

4   CONCLUSIONS

The concept of a real-time scale model auralization system applied to acoustic design provides a means for rapidly exploring tangible changes to the model, which can be heard immediately. Outside of acoustic design, such a tool allows one to use any miniature enclosure as an echo chamber for live performances or tangible reverberation customization for recording production, similar to early audio-production methods originating in the 1940s in recording studios[6]. This work presented the development of a numerical solution to achieve the goal of real-time scale model auralization, including a binaural receiver and spectral corrections for various source/receiver configurations. The SMALLROOMZ plugin is freely available at https://smallroomz.dalembert.upmc.fr/, providing instructions and several example recordings.

5   ACKNOWLEDGEMENTS

6   REFERENCES

1.   F. Spandöck, Raumakutsiche Modellversuche, Ann.Phys., 20(345), 1934.
2.   A. Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique, Proc. 108th Convention of the Audio Engineering Society, 1-23. Paris (2000).
3.   IRCAM LISTEN HRTF Database. http://recherche.ircam.fr/equipes/salles/listen/index.html. Last Access 31/07/2023.
4.   K. Spratt and J.S. Abel, All natural room enhancement, Proc. 2009 International Computer Music Conference, ICMC, 231-234. Montreal (2009).
5.   K. Delcourt, F. Zagala, A. Blum and BFG Katz. What's old is new again: using a physical scale model echo chamber as a real-time reverberator. Proc. 147th Audio Engineering Society International Convention, 1-11. New York (2019).
6.   P. Sutheim, An Afternoon with Bill Putnam. J Aud Eng Soc, 37 (9), pp.723–30, 1989.