Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, 3-5 April 2014

SCALE MODEL AURALIZATION FOR ART, SCIENCE, AND MUSIC: THE STUPAPHONIC EXPERIMENT

Brian F.G. Katz
Audio & Acoustics Group
LIMSI-CNRS
Orsay, France
`brian.katz@limsi.fr`

Markus Noisternig
Acoustic and Cognitive Spaces Group
UMR STMS, IRCAM-CNRS-UPMC
Paris, France
`Markus.Noisternig@ircam.fr`

Olivier Delarozière
Woodstacker
Champ-au-Beau, France
`olivier.delaroziere@woodstacker.net`

ABSTRACT

The use of acoustical scale models has been replaced for the most part by computational models and numerical simulations for room acoustic studies as well as artificial reverberation units. There remains however a number of acoustical phenomena which are difficult to address with computer simulations, such as coupled volumes, diffraction, and complex scattering, due to the computational complexity and/or calculation time necessary for addressing such acoustical wave phenomena on the scale of room acoustical problems, even small rooms. This paper presents a pilot study of a rather unique artistic architectural structure consisting of a self-supporting construction composed of small stacked linear elements. Acoustically, the structure combines modal behavior, concave forms, and very regular scattering patterns. An example scale model has been constructed and studied in order to separate different construction features and their associated acoustics effects. In an attempt to explore the interest of the specific acoustic for musical performance, a computational platform was created to utilize the scale model as a physical convolution reverberation unit for musical performance.

1. INTRODUCTION

With the advent of recording, and dry recording studios, there have been many efforts developed for the reintroduction of reverberation into studio recorded music. Some of the first technologies developed were the use of "echo chambers", wherein the dry audio captured with the microphone was diffused in a reverberant environment over loudspeakers, and then recaptured with microphones. This physical-based artificial reverberation was quite popular, with examples existing in such famous institutions as Abbey Road Studios where the echo chamber was constructed in 1931[a,b]. Echo chambers are, however, space demanding, difficult to transport, and not extremely adjustable. With improvements in electronics, other physical-electronic reverberation systems have been developed such a plate reverberators and spring reverberators.

With improvements in computer processing power, purely electronic reverberation became possible, such as using feedback delay network (FDN) for reverberation processing [1, 2, 3]. These reverberators could be easily adjusted, for example using perceptual descriptors relying on a simplified model of the time-frequency energy distribution of parametric FDN [4]. Such reverberators are however limited, lacking certain realism and ability to represent unique architectural elements. Additional increases in computational power allowed for the use of convolution reverberators, using complex impulse responses, either measured or calculated based on geometrical models such as ray tracing [5, 6], beam tracing [7, 8, 9], or radiosity [10, 11]. Convolution reverberators capture the fingerprint of a given space, but require preparations for the acquisition of such IRs and allows little flexibility with regards to modifying the room at time of use, though perceptual control of convolution based room simulators is a subject of current study [12].

Scale models, to date, have been used as off-line convolution reverberators to study architectural spaces [13, 14], but never in a performance setting. The current study envisages the possibility of using scale models in the same way as the old "echo chambers" of the early and mid-twentieth century, while allowing for the creation of complex and unique acoustic spaces rather than just simple reverberators. Real-time use of one, or several, models and the ability to dynamically alter source, receiver, and even room positions and configurations as desired offers a new form of reverberation and musical expression.

In parallel, the development and exploitation of real-time scale model convolution offers a number of interesting scientific aspects. To begin with, there is the basic signal processing challenge of achieving such a system. The applications of real-time physical-based convolution in scale models, in contrast to off-line convolution with measured impulse responses of the scale model, offers the ability to study room excitation by dynamic sources, such as moving or rotating, with perceptual studies. of specific interest are perceptual studies concerning musician/room interactions, which require real-time processing of generated music in coordination with source dynamics. The effect of dynamic architectures can also be examined, such as movable panels, or dynamic listener placement or movement during a performance.

---

[a] http://en.wikipedia.org/wiki/Echo_chamber, last viewed 2013-11-30

[b] http://audiogeekzine.com/2011/02/the-history-of-echo-echo-chambers-chambers/, last viewed 2013-11-30

2. ARTISTIC CONTEXT

What child has not dreamed of being able to experience, as a Lilliputian, a world in miniature: doll houses, electric trains, miniature circus . . . In the "Stupaphonic" [1] project that childhood dream will become a reality for musicians: they will be able to play to their audience in a space in miniature.

This particular space is based on a special type of structure, which is at the core of the architectural project *Woodstacker* [15]. This architectonical type of structure is a solution to the geometrical problem of how to cover a large area by stacking small pieces of wood without the use of glue or nails. The result is a bottle shaped three-dimensional rose window (see Figures 1 and 4). This new building system, based on "stacked laminated" timber structures, can evoke references to *pagodas* whose construction also consists of wooden stacked elements. This stacked architecture, like *chorten* of Tibet, belongs to a family of stacked structures which are derived from a Buddhist mound-like structure called *stupa*. Stupas originated as pre-Buddhist earthen burial mounds, like *tumuli* in Europa. Thus was born the idea of linking our new project to these ancient and universal architecture.

The stupa is used by Buddhists as a place of meditation. In the original pre-Buddhist burial mounts ascetics were buried in a seated position. The American anthropologist J. Jaynes [16] proposes that they were buried in this position so they can continue to speak to living people. Hearing voices from beyond the grave suggests some acoustic illusions which are also a part of our device. In the Woodstacker system, the special geometrical pattern of the lamellas not only focuses the sound [17] but also functions as frequency filters producing a particular, almost metallic, sound. This strange acoustic effect is a second reason to link our project to the *stupa* as a kind of container for "voices from beyond the grave", a "voice granary" ("grenier à voix" in French) to cite the french writer Pascal Quignard [18].

The larger structures we have currently built can accommodate about 30 people for sound experiments (see photograph in Figure 1). This size limitation is a compromise between the funding for artistic experiments and the cost of such a construction. With the project's evolution we desired a means to quickly experiment with different architectural structures in a flexible way. The use of physical scale models which are powerful tools for architects, carpenters, and acousticians [19] offered a solution to exceed the current constraints. Thus, we found a way to invert the acoustical environment, like turning a glove inside-out, and give the musician and audience located outside of the building the same acoustical experience that they could have inside the structure. Using the acoustic scale effect we are able to drastically reduce the size of our installation and increase the number of structures performers can play with and turn "stackscapes" (see Figure 2) into interactive soundscapes.

3. SYSTEM OVERVIEW

To achieve the artistic, acoustic, and audio scheme imagined, a basic scenario and system architecture was envisioned. One can imagine a performance area, where the performer is equipped with one or several microphones. The space is open and large. Near the musician is one or several acoustic scale models, equipped with ultrasonic speakers and microphones. Around the musician is the audience. The sound produced by the musician is captured, transformed to the scale of the model where it is played and recaptured, then transformed to the full scale of the musician's performance, where it is played live to the audience over an electro-acoustic array of speakers either on-stage or around the audience.

Figure 1: *Photo of live performance at StackCamp 2013 featuring Didier Petit (cello) and Emre Gultekin (saz). Champ-au-Beau, France.*

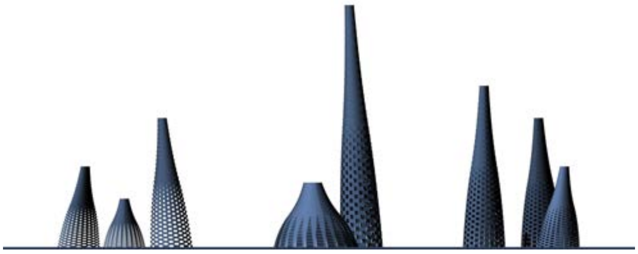

Figure 2: *Blue Stackscape, ©2006 O. Delarozière.*

3.1. Signal Processing

The signal processing chain is depicted in Figure 3. First, the input audio signal is up-sampled to the ultrasound sampling rate, which is determined by the scaling factor of the model. Second, the up-sampled input signal is transposed by the scaling factor preserving the harmonic structure of the signal. Two implementations have been tested: a) off-line transposition that allows for the time-stretching of the signal; b) a real-time implementation thereof, which compensates for the time-scaling effect using

---

[1] Stupaphonic: from stupa (from Sanskrit: m., स्तूप, *stūpa*, literally meaning "heap" [a]) and phonic (from Ancient Greek *φωνή*, *phōnē*, meaning "voice" or "sound" [b])

---

[a] http://en.wikipedia.org/wiki/Stupa, last viewed 2013-11-30

[b] http://en.wikipedia.org/wiki/Phonetic, last viewed 2013-11-30

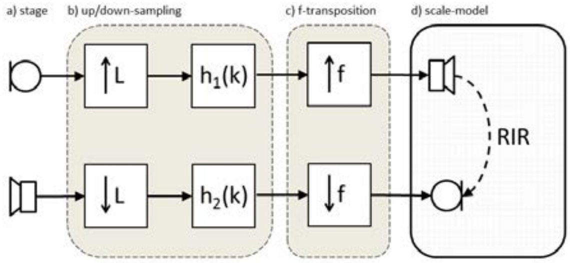

Figure 3: *Signal processing flow chart: (a) instrument signal capture, transformed audio signal playback; (b) up/down-sampling with anti-aliasing filters; (c) frequency transposition, and (d) the Stupaphonic scale model.*

a phase vocoder and thus preserves the continuity of the audio stream.

The off-line version study was carried out using MATLAB®. Audio samples were processed individually, with no audio streaming functionality. The basic approach consisted of taking an audio extract, resampling the audio, modifying the sample rate in order to apply the scale factor, play/recording the sample in the model, and then retransforming the recorded physically-convolved signal to full-scale for audio playback. A code sample for the described process is provided below:

```
fs = 44100;
[y] = wavrecord(audiolength,fs);
fs_max = 192000; % sample rate in scale model audio chain
scale_desired = 12;
[y_resamp] = resample(y,fs_max/scale_desired,fs);
fs_resamp = fs_max/scale_desired;
[y_convolved]=wavplayrecord(y_resamp,fs_max);
wavplay(y_convolved,fs_resamp);
```

In this example, with a maximum sample rate of 192 kHz on the audio system and a scale factor of 12, the recorded audio track is resampled to 16 kHz (bandlimited to 8 kHz). This resampled track is then played back and recorded in the scale model at an actual sample rate of 192 kHz. The recorded physically-convolved signal is then played back at an actual sample rate of 16 kHz, or resampled to the sample rate of the audio device. The simple redefinition of the sample rate for the audio buffer performs the application of the scale factor, while the resampling assures correct anti-aliasing filters.

While currently tested in single buffer full convolution, future studies will evaluate the possibility of applying the concept of overlap-add convolution [20] to the concept of this physical-based convolution in order to allow for real-time operation on audio streams.

The real-time version study was conceived of as an alternate approach to the above approach employing resampling and transposition to apply the scaling factor through the use of a high quality phase vocoder architecture. Phase vocoder techniques are typically based on a sinusoidal signal model. The digital audio sampling rate conversion employed band-limited interpolation (see e.g. [21, 22]) that can be efficiently implemented with sinc-function look-up tables. In [23] it was shown that parametrized phase vocoders can also be applied to non-sinusoidal signals. However, initial tests showed that the sinusoidal signal model limits the use of phase vocoders for real-time scale-model processing for large scale factors. Modified algorithms are the subject of continuing investigations.

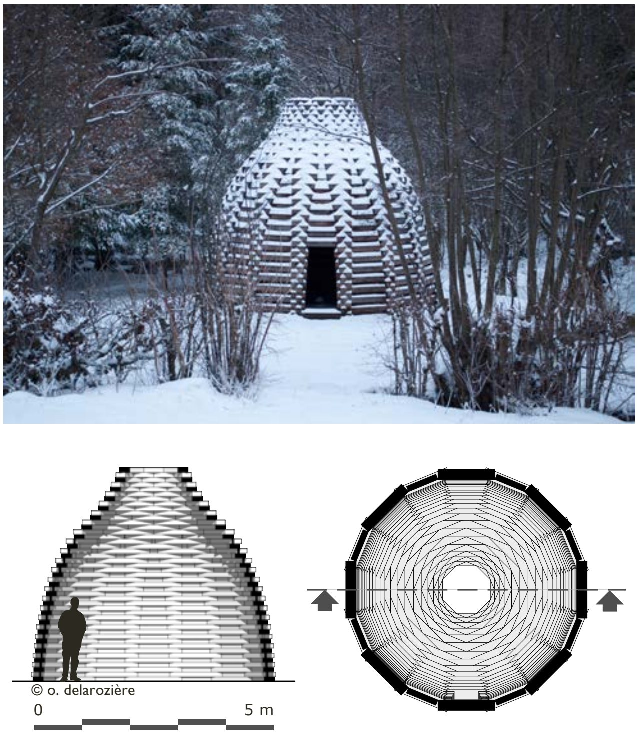

Figure 4: *Woodstacker stacked lamella timber cupola. Champ-au-Beau, France, 2010. (upper) A winter outside view. (lower) Section and Reflected ceiling plan.*

3.2. Scale Model

This preliminary study has been carried out using a single structure as a test case. The scale 1:1 structure was built in 2008 for a Land Art Exhibition in the highland of Auvergne, France (see Figure 4). The original building comprised 366 pieces of Douglas pine wood. It was 5 m in diameter, 4.5 m high, and weighed 6 tons. This work called “Vox Granarium” [24] was dedicated to famous ancient fiddlers from this area. This installation was dismantled and moved to Morvan where it was rebuilt in 2010. The Stupaphonic model is a 1:12 scale model of “Vox Granarium”, 425 mm in diameter and 322 mm high. This scale was chosen due to material availability and because it is a traditional doll house scale. Serendipitously, 1:12 is also the Lilliputian people’s scale in Gulliver’s Travels[2]. The model was constructed from oak wood, whose lengths and widths were hand cut with no automatic process used for assembly. Unlike the full-scale construction, glue was used to fix the lamellas, and the model was assembled in three parts for ease of transportation and manipulation purpose (see Figure 5).

Due to the very long time needed to build this model by hand, subsequent models will probably be built using rapid prototyping techniques such as laser cutting. This will allow to quickly experiment with a large variety of structural shapes and configurations. We are also planing to use other materials such as metal or concrete for special acoustics effects.

---

[2] “. . . having taken the height of my Body by the help of a *Quadrant*, and finding it to exceed theirs in the Proportion of twelve to one . . .” [25, p. 64]

Figure 5: *Scale model (1:12) of "Vox Granarium", highlighting the tetrahedral acoustic source and 3 modular elements.*

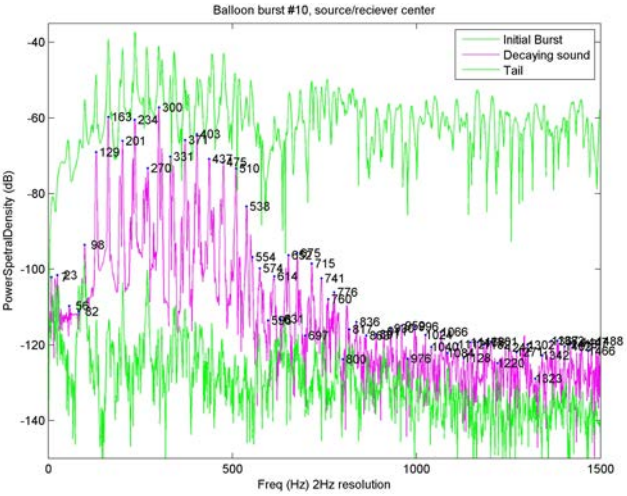

Figure 6: *Spectral analysis of impulse response (balloon burst excitation) of the full (1:1) scale "Vox Granariium", recorded with a sample rate of 96 kHz. Temporal analysis windows corresponding to direct/initial* ($0-10$ *msec), early decay* ($10-30$ *msec), and late tail* ($30-\infty$ *msec).*

4. PRELIMINARY TEST

The diffraction acoustic effect discussed in Section 2 can be observed as a series of high-Q total resonances. These resonances can be observed by comparing the spectral response (magnitude of the FFT) at different moments in the impulse response. Figure 6 shows the spectral response at three moments in the impulse response (balloon burst excitation signal employed for room acoustics measurements [26]) of the full (1:1) scale "Vox Granariium" (see Figure 4). Resonant peaks are identified in the response for identification. There is clearly a region of resonance density over the frequency range 100–600 Hz, continuing still to ≈ 800 Hz.

The audio system employed for use in the scale model consisted of DPA 4060 microphones and a custom 3-speaker tetrahedron (see Figure 5) driven by a Samson Servo amplifier, connected to a RME Fireface 400 audio interface. While somewhat unconventional in traditional scale model research, this selection of pro-audio equipment has been used in previous studies [27, 28] and has been shown to provide improved signal-to-noise ratio when compared to more traditional laboratory scale model measurement architectures. The current hardware exhibits a frequency roll-off at ≈ 50 kHz. This of course imposes a low-pass frequency limitation for the physical-based convolution. With a scale factor of 12, the upper frequency limit due to this roll off is on the order of 4.2 kHz, rather than the 8 kHz permissible due to sampling theory. While suitable for the majority of studies in room acoustics with scale models, the musical implications of this limit can not be ignored. This limit can be raised by improving the upper frequency limit of the audio chain or selecting a lower scale factor.

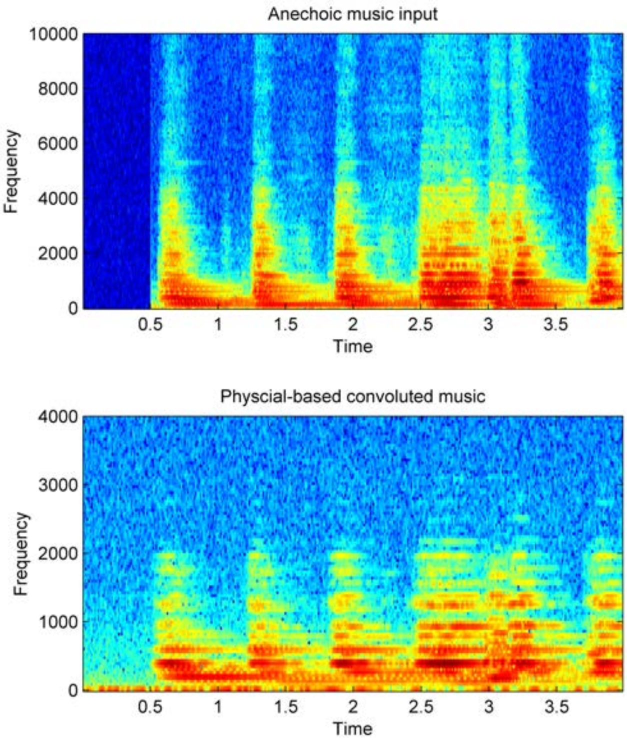

Figure 7: *Spectrogram of anechoic music extract (upper) and physical-based convolved music (lower) manually time-aligned.*

An example result of the processing chain can be seen in Figure 7, which shows the spectrogram of a dry music extract before and after the physical-based convolution processing. The test music except was a dry multichannel recording of a Schubert trio (D.929, op.100), by [29] and publicly available [a].

The acoustic timbre of the convolved signal using the scale model greatly resembled that of the musical experience heard within the full scale installation. Even thought the processing steps apply a low-pass filter effect, due predominantly to transducer and amplifier performance limitations above 50 kHz, the frequency range where the resonance characteristics of the structure are apparent are still well within the operating frequency of the current signal processing chain for the 1:12 scale.

5. CONCLUSIONS

This paper has presented the foundations of the "Stupaphonic experiment", an artistic and scientific project which aims to use scale models as physical-based convolution reverberators. The architectural structure at the center of the project offers specific timbral qualities which are maintained in the initial tests, despite the frequency limitations of the scale transformations and associated signal processing chain.

The current example operates in an off-line, or time-deferred situation. While streaming is currently still being investigated, the current implementation could still be used in a performance setting

---

[a] http://c4dm.eecs.qmul.ac.uk/rdr/handle/123456789/27

in a live looping context, where the musician could send different audio samples to different architectures at a unique or different scale factors, effectively changing the size of the "echo chamber".

The development of the real-time processing stage, currently a subject of study, will allow for exploitation of the proposed physical-based convolution for studies in room acoustics, specifically those involving dynamic source, listener, or architectural elements, as well as dynamic performer/room interactions.

One artistic performance aspect specific to this system is the potential for cross-scale cross-talk. If the scale models are open to some degree, then the up-scaled audio will be heard by some of the audience. At the same time, full scale sounds, such as other elements of the performance or noise from the audience, can also be captured in the scale model, and subsequently down-scaled and played over the reproduction array. According to the Lilliputian scale factor of 1:12, one can imagine the majority of these sounds will be shifted to the lower end of the audible range, or into the subsonic region. However, a high pitched scream with a center frequency in the 5 kHz third-octave band for example would be clearly audible when transposed to the 400 Hz third-octave band, albeit also time stretched to 12 times its original duration. Investigations of these effects, and their possible artistic use, remain the subject of further studies.

6. ACKNOWLEDGMENTS

This study was funded in part by an Action Initiative grant at the LIMSI-CNRS.

7. REFERENCES

[1] J-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Proc. 90th AES Convention*, Paris, France, Feb. 1991.

[2] G. Garcia, "Optimal Filter Partition for Efficient Convolution with Short Input/Output Delay," in *Proc. 113th AES Convention*, Oct. 2002.

[3] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3D Ambisonic based Binaural Sound Reproduction System," in *Proc. 24th AES Int. Conf.*, Banff, Canada, June 2003.

[4] J-M. Jot, "Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces," *Multimedia Systems*, vol. 7, no. 1, pp. 55–69, 1999.

[5] A. Krokstad, S. Strom, and S. Sorsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound Vib.*, vol. 8, no. 1, pp. 118–125, 1968.

[6] M. R. Schroeder, "Digital Simulation of Sound Transmission in Reverberant Spaces," *J. Acoust. Soc. Am.*, vol. 47, no. 2, pp. 424–431, 1970.

[7] T. A. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. West, "A beam tracing approach to acoustic modeling for interactive virtual environments," *Proc. ACM Comp. Graphics (SIGGRAPH'98)*, pp. 21–32, July 1998.

[8] S. Laine, S. Siltanen, T. Lokki, and L. Savioja, "Accelerated beam tracing algorithm," *Applied Acoustics*, vol. 70, no. 1, pp. 172–181, 2009.

[9] M. Noisternig, B. F.G. Katz, S. Siltanen, and L. Savioja, "Framework for real-time auralization in architectural acoustics," *Acta Acoust. united with Acust.*, vol. 94, pp. 1000 – 1015, 2008.

[10] C. Malcurt, *Simulations informatiques pour prédire les critères de qualification acoustique des salles. Comparaison des valeurs mesurées et calculées dans une salle à acoustique variable*, Ph.D. thesis, Laboratoire Acoustique Métrologie Instrumentation, Toulouse, France, July 1986.

[11] G. I. Koutsouris, J. Brunskog, C-H. Jeong, and F. Jacobsen, "Combination of acoustical radiosity and the image source method," *J. Acoust. Soc. Am.*, vol. 133, no. 6, pp. 3963–3974, 2013.

[12] T. Carpentier, T. Szpruch, M. Noisternig, and O. Warusfel, "Parametric control of convolution based room simulators," in *Proc. Int. Symp. on Room Acoust. (ISRA)*, Toronto, Canada, June 2013.

[13] Jean-Dominique Polack, Xavier Meynial, and Vincent Grillon, "Auralization in scale models: Processing of impulse response," *J. Audio Eng. Soc*, vol. 41, no. 11, pp. 939–945, 1993.

[14] Vincent Grillon, Xavier Meynial, and Jean-Dominique Polack, "Auralization in small-scale models: Extending the frequency bandwidth," in *Audio Engineering Society Convention 98*, Feb 1995.

[15] O. Delarozière and U. Gleeson, "Woodstacker," in *Architectures autrement : Habiter le monde*, M. Culot and A-M. Pirlot, Eds., pp. 46–51. AAM, Brussels, 2005.

[16] J. Jaynes, *La naissance de la conscience dans l'effondrement de l'esprit*, Presses Universitaires de France, 1994.

[17] B. Katz, O. Delarozière, and P. Luizard, "A ceiling case study inspired by an historical scale model.," in *Proc. 8th Int. Conf. on Auditorium Acoust., Institute of Acoustics*, Dublin, May 2011, vol. 33, pp. 314–321.

[18] P. Quignard and C. Lapeyre-Desmaison, *Pascal Quignard le solitaire : Pascal Quignard, rencontre avec Chantal Lapeyre-Desmaison*, Les Flohic éditions, 2006.

[19] O. Delarozière, "Camera tectonica : Hypothèses pour un facsimilé d'architecture," in *Utopia Instrumentalis : Facsimilés au musée - Musée de la Musique*, Cité de la musique, Paris, Nov. 2010, pp. 46–56.

[20] A. V. Oppenheim and R. W. Schafer, *Digital signal processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975, ISBN 0-13-214635-5.

[21] R. W. Schafer and L. R. Rabiner, "A digital signal processing approach to interpolation," *Proceedings of the IEEE*, vol. 61, no. 6, pp. 692–702, 1973.

[22] J. O. Smith, III and P. Gossett, "A flexible sampling-rate conversion method," in *Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proc. (ICASSP)*, 1984, pp. 112–115.

[23] W.-H. Liao, A. Roebel, and A. W. Y. Su, "On stretching gaussian noises with the phase vocoder," in *Proc. of the 15 Int. Conference on Digital Audio Effects (DAFx-12*, Sept. 2012.

[24] O. Delarozière, "Vox Granarium," in *Horizons - Rencontres "Arts Nature"*, pp. 12–13. Office de Tourisme du Sancy, July 2008.

[25] J. Swift, *Part 1. A Voyage to Lilliput*, vol. 1 of *Travels Into Several Remote Nations of the World*, chapter III, pp. 47–64, Printed for *Benj. Motte*, at the Middle *Temple-Gate* in *Fleet-street*, 1726.

[26] J. Pätynen, B.F.G. Katz, and T. Lokki, "Investigations on the balloon as an impulse source," *J. Acoust. Soc. Am.*, vol. 129(1), pp. EL27–EL33, 2011.

[27] Paul Luizard, *Les volumes couplés : comportement, conception, et perception dans un contexte de salle de spectacle*, Ph.D. thesis, Université Pierre et Marie Curie, Paris, France, 2013.

[28] P. Luizard, M. Otani, J. Botts, L. Savioja, and Brian F. Katz, "Comparison of sound field measurements and predictions in coupled volumes between numerical methods and scale model measurements," in *Proc. Meetings on Acoustics*, Montreal, June 2013, vol. 19, p. (9 pages).

[29] Joachim Fritsch, "High quality musical audio source separation," M.S. thesis, UPMC / IRCAM / Telecom ParisTech, 2012.