國立台灣大學電機資訊學院電機工程研究所

碩士論文

Department of Graduate Institute of Electrical Engineering
College of Electrical Engineering and Computer Science
National Taiwan University

Master Thesis

水下聲源於玻璃缸中之聲學模擬與其在動畫配音之應用

Auralisation of Under Water Sound Sources in Glass Containers with Applications in Animation Sound Designing

羅晟文

Sheng-Wen Lo

指導教授：鄭士康 博士

Advisor: Shyh-Kang Jeng, Ph.D.

中華民國 101 年 6 月

June, 2012

致謝

Auguste Rodin 曾說 “Nothing is a waste of time if you use the experience wisely.” 現在，我可以開心的說，在碩二下時加入鄭教授的電腦音樂實驗室，是我碩班期間做過最好的決定；這三個學期裡，我可以自由的做自己喜歡的研究，和喜歡的事。

非常感謝鄭教授一年半來的指導。記得剛加入實驗室時，教授告訴我，我的研究主題和進度都是我自己的選擇，也是自己的責任。從那時起，我才了解什麼叫自己對自己負責，以及如何進行獨立研究。找尋研究方向期間，每次報告後，不管我的問題是多麼的基本，進度是多麼的緩慢，教授總是能耐心的為我分析狀況，並精準的告訴我問題所在與各種可行的方向。更令我敬佩的是，我感受到鄭教授對科學研究產生的道德問題非常重視，這對我來說非常重要，以此為科學研究應有的初衷，讓我十分感動。

感謝韋安在我狀況最差時介紹我加入實驗室，並邀我一起修習聲景與聲音藝術課程，開拓了我對聽覺世界的了解，生活中的細節從此也更加有趣了。修課期間和韋安、伊里一起想問題，一起創作是我最快樂的時光，我們合作的《SDINGUHOTS》錄像裝置是影響我生活態度最重要的作品，其精神也會一直延續下去。感謝伊里幫忙我剪輯這篇研究中產生的聲音片段，並每天和我想著各種大小事，偶爾甚至在實驗室做作品做到天亮；利用攝影，機械和電腦音樂知識做出的《萬花筒計畫 I 》是我最愛的互動聲音裝置。感謝志鴻學長天天打點著實驗室裡外各種事，以及常對我的研究 (或是和研究一點關係也沒有的事物) 提出有趣的疑問和點子；感謝鴻欣學長主辦週末電影院，不管選到好片或爛片，燈一關，大家總是很歡樂。謝謝御仁，彥彬學長以及 JCMG 的大家，在報告後提出許多我沒想過的好問題。感謝實驗室，提供輕鬆的空間與齊全的軟硬體設施。更要感謝家人的支持，教育與努力，讓我在學期間可以安穩的做著研究與創作。

感謝審查委員：曾毓忠教授與鍾世凱教授，細心的閱讀我的論文，口試時指點了必要的加強之處；感謝黃佳惠學姊，在我口試期間給了我好多的幫忙。謝謝吃飯團與攝影社的朋友們給我許多關心與支持；也一定要謝謝令網，筱倫與 Rajkumar Hirani 導演的《3 idiots》，因為你們給了我轉換實驗室的思想基礎與勇氣！

在台大轉眼間過了七年，每一年都很精采，有開心的精采，也有難過的精采。每次年底時都會覺得這應該是最精采的一年了，但下一年總是又以另我驚訝的姿態展開。現在，我每一刻都感受著這七年帶給我在思想上的影響，也逐漸認識著，探索著自己：我了解到如何不去限制自己，也知道我喜愛令我驚奇的事物，謝謝大家給了我七年的驚奇，不用說再見，畢業僅是另一個開始！

中文摘要

聲學模擬科技自上世紀以來已有長足之發展；然而，鮮少有人將這類技術應用到藝術與創意相關領域上。理論上，許多與聲音相關的產業與藝術創作都有可能受惠於聲學模擬技術。本研究專注於將聲學模擬科技應用於動畫配音上，並主要探討兩種聲學模擬科技：指數正弦掃頻法與有限元素分析法。本研究中，此兩方法被用於模擬一個想像的狀況——聲源沉浸於盛滿水之玻璃缸中，聽者自缸外聽之。雖然此兩聲學模擬科技有其限制，本研究將指出並展示它們在藝術與創意領域中的應用潛力，說明它有機會為聲音藝術開拓新的視野。

關鍵字：聲學模擬、動畫配音、指數正弦掃頻、有限元素分析法、水下聲源

Abstract

The science of sound simulation, or auralisation, has been developing fondly for decades; however, little effort has been made to utilize such techniques in the art and creative industries. Theoretically, many sound related creative workflows may benefit immensely from the technologies of auralisation. This thesis focuses on applying such technologies in the field of sound designing of animations. Two methods, namely the exponential sinesweep method and the finite element method, have been implemented on a specific scenario – a sound source placed in a water-filled glass container, heard by the listener at the outside. It can be shown and demonstrated by this thesis that although these modern techniques are not without limitations, they have great potentials to broaden the horizon of sound designing or other sound related art and creative fields.

Keywords: auralisation, animation sound design, exponential sinesweep, finite element method, underwater sound sources

CONTENTS

LIST OF FIGURES

Chapter 1
Introduction

1.1 Motivation and Objectives

Over the past decades, acoustic engineering and digital sound technologies have been aiding the creative processes in numerous fields, such as animation, music composition and live performances. The potential application targeted in this thesis is the sound designing of animations and films. Nowadays, advanced editing software and well-developed sound banks provide sound artists with sped up workflows and pleasing results. However, little research has been done on the most basic aspect of sound designing: the generation of sound itself. In fact, modern sound designers rely heavily on studio recorded sounds and actual field recordings to acquire their raw materials at the core of the creative process.

Oftentimes, the scenarios which animators or filmmakers imagine are impossible or impractical to realize in the real world. In this case, sound designers and foley artists work their magic by relating real life recordings with the imagined scenarios. This process relies immensely on the experience and virtuosity of sound professionals, and is considered artistic in nature. Undoubtedly, the result they have achieved, as we have seen in lots of modern films and animations, are nothing short of brilliant. However, there are also times that compromises have to be made, and consequently the question can be raised: can auralisation technologies aid this designing process, and make the results even better? Fortunately, thanks to the advancements of computing and signal processing techniques, we are given an opportunity to reconsider the possibilities of sound generation. In order to further expand the horizon of the creative world, related technologies and their potential applications should not be overlooked. For instance, the acoustic technologies developed in concert hall structural designs may have direct or indirect applications in the above-mentioned creative fields. We shall discover that some techniques are as readily available in these creative fields as they are in the architectural world. Furthermore, with the rapid development of hardware computing power, computational intensive numerical methods such as the finite element method may also be taken into account. The key advantage of these methods is that they do not need any actual field recordings to produce sounds; they rely purely on 2D or 3D models and physical rules to perform calculations of acoustic simulations. Hence, such technology may offer the creative fields a chance to create sounds that are conceivable in the designers' or the composers' minds but not recordable in the actual world.

In this research, a scenario in which a sound source is placed in a water-filled glass container is conceived. The size and materials of this scenario is set to be challenging for modern auralisation techniques. Two methods, namely the exponential sinesweep method and the finite element method, are utilized in attempt to auralise this scenario.

The objective of this thesis is therefore stated below:

- **Explore and demonstrate the possibilities of applying auralisation technologies, such as the exponential sinesweep method and the finite element method, to the creative fields, such as sound designing for animations and films.**

1.2 Literature Survey

There has not been many related research previously conducted on applying acoustic science explored by this research to the creative fields. Nevertheless, auralisation in buildings and rooms such as concert halls is the closely related to the field of interest; a review is given below.

Auralisation in Architecture

In analogy with the term "visualization", the term "auralisation" was coined to describe the process and the result of rendering audible sound fields. A more detailed definition was given by [1], stated in appendix A.

There has always been a need for architects of opera houses and concert halls to "hear" in advance how their designs would sound like before they are actually built. Therefore, building acoustics has been applying auralisation technologies for more than half a century. Back in the 1930s, Spandöck [2] and his research team uses ultrasonic techniques and scaled building models with special speakers placed in them. The sound was recorded in analog fashion, and the recording speed and playback speed are manipulated meticulously to produce an approximation of an auralised listening experience. We can call this kind of analog technique "Acoustic Scale-Model Auralisation". Similar methodologies can be found in Japanese researches, such as Oguchi et al. [3].

With the advancement of sound recording, digital processing, and computing technologies, “Fully Computed Auralisation” has become the main breed of technologies used by architectural acoustic simulation. The term “fully computed” means no physical model has to be built at all. These comprise three methodologies: the ray tracing method, the image source method, and the numerical methods. Generally, in these methods, a 3D digital model of the room or building is first constructed. Then, physical rules are applied to represent the behavior of sound waves, and the subsequent calculations are based on 3D models. A brief survey and description is given below.

Fig. 1-1 Ray tracing method [1]

- **Ray Tracing Method**

  This method models the sound waves between the sound source and the listener as a large number of very narrow beams (rays). The propagation time and the decay of the rays are calculated as they reflect and diffuse against walls. Only a small number of rays radiate from the source may contribute to the listener; thus, tracing only these contributive rays can speed up calculation. “Odeon A/S”, a modern architectural auralisation software, has been using ray tracing method since the time it was developed.

- **Image Source Method**

As the name suggests, this method takes advantage of the similarity between sound sources and image sources. For instance, walls can reflect sound waves, so they are considered “mirrors” for sound sources. The resolution or the efficiency of the mirror effect depends on the walls’ surface geometry and materials. The image source method suffers from some known problems and has to be corrected. The Odeon A/S software now utilizes both the ray tracing method and the image source method to correct each other, in attempt to gain the advantage of both methods.

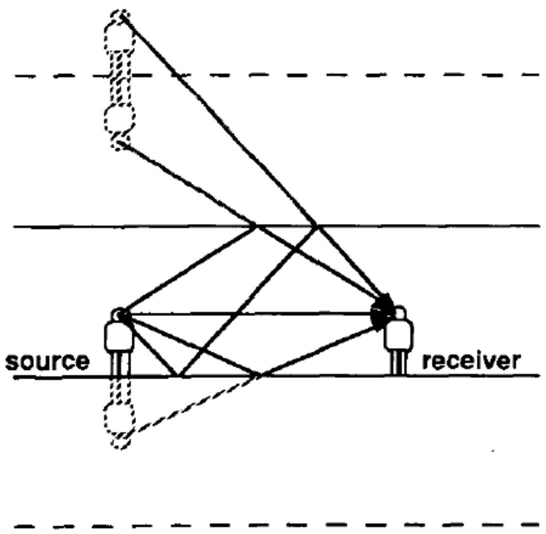


Fig. 1-2 Image source method [1]

- **Numerical Methods**

The two methods introduced above require a lot of assumptions, and does not account for many wave-related phenomena such as scattering and diffraction. For those methods, the size of the space of interest has to be large, or the assumptions will fail. In other words, the two methods and their hybrid applications like Odeon A/S works well for large scale buildings such as opera houses and churches, but would not be accurate for smaller room or spaces like recording studios and home theaters. Numerical methods such as the finite element method (FEM) account for these phenomenon by modeling the entire 3D model with tiny adjacent fragments of space. That is to say, FEM approximates 3D models with tiny fragments and calculate the effect the waves contribute to each fragment. Theoretically, summing up the effects yields accurate result if the fragments are “small enough”. However, having had to account for all the fragments can be extremely computationally intensive; thus, in the old days, the FEM was often considered impractical. Thanks to the rapid development of hardware computing power, FEM has been used more often in in small room acoustics in the past two decades. In these cases, “small rooms” refers to the rooms which dimensions are in the same order with the sound wave frequency of interest. In the 1990s, Craggs [4]

[5] conducted researches on modal analysis on rectangular, triangular as well as cylindrical small rooms with absorptive patches and linings based on FEM. The work focused on the modal analysis and absorptive material performance, and no actual auralisation was done. In 2008, Sakamoto et al. [6] attempted to calculate the acoustic characteristics (the impulse response) and the acoustic parameters of a large hall using finite-difference time-domain (FDTD) method, which is another numerical method. Comparing with the actual response, the results are decent within middle-frequency bands, but large discrepancies were observed in the low-frequency band. The limit of the operational frequency range of FDTD further limits the result. Similarly, no actual auralisation was presented in [6]. Acoustic analysis of a small room carried out by the FEM software COMSOL multiphysics can be found in [7]; still, no actual auralisation was done. In this research, a peculiar case involving water-filled glass containers is considered. Physically and geometrically, the major differences between this case and the prior researches are that the glass containers are not enclosures and are way smaller than small rooms. Since this is an attempt on actual auralisation of the scenario, time domain analysis is of major concern.

1.3 Key Contributions

The Key contributions of this research are stated below:

- Explore the possibilities of applying modern auralisation methods to water-filled glass containers.
- Provide a new methodology of rendering realistic sound tracks base on impulse responses.
- Retrieve the acoustic characteristic of a water-filled glass container by applying the exponential sinesweep method.
- Demonstrate and discuss the advantages of utilizing auralisation techniques on animated films.
- Experiment the possibilities and limits of applying finite element method on auralising the target scenario.

1.4 Chapter Outline

:

In chapter 2, a review of background knowledge will be given, focusing mainly on the related fields of sound designing in films and animations and the characteristics of linear systems. These backgrounds provide the knowledge on which the following research is based. In chapter 3, the target scenario and the physical parameters of the glass tanks are given. In chapter 4 and 5, the theories and experiments of this work are explained and carried out, and the physical and software-based model of the scenarios are made. Two methods, namely the exponential sinesweep method and the finite element method, are adopted in attempt to auralise the target scenario. In chapter 6, the results are presented, and the discussions for experimental successes as well as failures are carried out. For demonstration purpose, the achievement of the experiments of the previous chapters is applied to a selected animation clip. Eventually, in chapter 7, the conclusion is drawn and the future possibilities of this study are noted.

Chapter 2
Background Knowledge

Since the application of the thesis focuses mainly on sound designing in animations and films, an overview of the designing process will first be given. The references of the following section can be found in [8] and [9], and the reader is encouraged to explore these background materials.

2.1 Sound Designing in Animations and Films

Creating sounds for animations is a challenging task since the whole movie essentially requires an original new world of sound. Sound designers begin with figuring out the overall language and atmosphere of the animated film, and set the overall soundscape on that basis. Sound design has a long history, dated way back into the 1920s when the digital tools are not available at all. For example, Disney sound artists back then had the ability to create rich and accurate soundscapes for Mickey Mouse movies such as “The Karnival Kid” (1929). Their classical style is to use musical instruments to produce sounds to accompany visual effects like collisions, explosions and horrendous weathers. Later on, with the creative minds of sound artists such as Jimmy McDonald, “sound devices” begin to flourish in the industry. These sound devices are machines custom built to create all sorts of sounds. For instance, a “rain machine” features a rotating metal cylinder that can go fast and slow with Mexican beans tumbling along the rough inner surface; a “wind machine” has a similar rotating structure that can mimic the sound of gusts of winds by scraping the canvas wrapped around bars made of woods; a “thunder sheet” is simply a large piece of sheet metal which when vibrated can generate “a wonderful low frequency rumble out of it”, commented Ben Burtt, the father of modern sound design who created the soundscapes of films including Star Wars (1977), Indiana Jones (1984) and Wall-E (2008).

In spite of the advancement of computer technology, the fact is that the basic aspects of sound design have not change much since the introduction of sound devices. For instance, according to Ben Burtt, the “wind machine” and the “thunder sheet” are utilized in modern motion pictures such as Wall-E. Though digital sound banks are available for purchase, seasoned modern sound designer and foley artist still walks around the actual world with their recording devices and their ears wide open. In other words, they create their own unique sound libraries by gathering sounds in a documentary style, believing that they must be useful someday. The difference between modern and classic sound design lies in post processing. Thanks to the aid of digital post processing software, the final renderings created from these raw recorded materials are more sensational than ever. In studios, sound designers can now orchestrate the recordings together to create powerful results that can be convincing enough to punctuate emotions.

Another technique worthy of notice is how the sound designers create a sense and atmosphere of special space in which words of a character is spoken. In this scenario, designers may either build or find an appropriate space, and play the pre-recorded vocal lines in it. The resulting sound is then recorded for further post processing. Notable examples can be found in the Lord of the Rings movies, where voices of “Ents” and “cave trolls” are played and recorded in had crafted wood chambers and an old tunnel in Willington respectively. Despite the fact that the final sound track works well, this method raises some concerns on the reusability and accessibility of “spaces”. In this thesis, this particular problem is addressed with the use of signal processing. This research will reveal that utilizing such techniques may allow foley artists to produce a whole new kind of reusable “sound bank” – the acoustic responses of all sorts of spaces of interest.

2.2 Backgrounds on signal processing

Since the theories on which this research bases lies with the relationship between linear time-invariant systems and their impulse responses, a brief background is presented here. This research performs digital signal processing; therefore, only discrete-time signals are considered here. For a more complete introduction including continuous time systems, the readers are recommended to refer to the first two chapters of [10]. The figures for illustration of this section are adopted from [10] as well.

2.2.1 Discrete-time signal

When a digital recorder records a sound, it is actually taking successive samples of the intensity of air vibration. That is to say, the analog sound is recorded as a sequence of numbers, ordered by time. This sequence of numbers is a discrete-time signal, defined only at discrete times. We can denote the values of a discrete-time signal as $x[n]$, where n takes only integer values. Figure 2-1 is a graphical representation of a discrete-time signal.

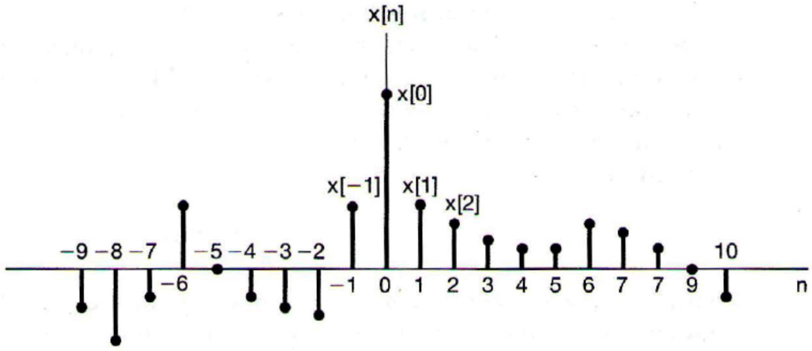


Fig. 2-1　An example of a discrete-time signal, excerpt from [10]

2.2.2　Unit impulse

Here we consider a special but simple discrete time signal: the discrete-time unit impulse. It is defined as:

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 1 \end{cases} \tag{2.1}$$

An illustration is shown below:

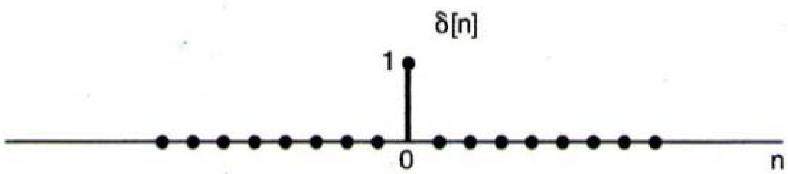


Fig. 2-2　Discrete-time unit impulse, excerpt from [10]

The unit impulse is the basic building block for the construction and representation of other signals. In fact, any discrete time signals $x[n]$ can be represented by a summation of shifted and scaled unit impulse; one may represent this fact as below:

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\,\delta[n-k] \tag{2.2}$$

2.2.3 Linear Time-invariant System

A system in a broad sense can be regarded as a process that can take the input of some sort, respond to the input in some way, and produce the output base on the response. For example, a digital music player takes the digitally sampled sound signals and converts it into analog music as the output.

Fig. 2-3 Discrete-time system

A discrete-time system is a system that takes some discrete-time input and produces the corresponding discrete-time output. It is illustrated in Fig. 2-3, and can be expressed symbolically as

$$x[n] \rightarrow y[n] \tag{2.3}$$

Here, two important properties of a system are introduced: linearity and time invariance. A system is *linear* if it possesses the property of superposition. That is, if the inputs of the system are a weighted combination of signals, the output is the weighted sum (superposition) of the outputs produced by each of these signals. By (2.3), let $y_1[n]$ be the response of a discrete time system with the input $x_1[n]$, and $y_2[n]$ corresponds to the input $x_2[n]$, then the system is linear if:

1. The response to $x_1[n] + x_2[n]$ is $y_1[n] + y_2[n]$. (Additive)
2. The response to $kx_1[n]$ is $ky_1[n]$, where k is any complex constant. (Scalable)

A system is said to be *time invariant* if its behavior does not change over time. In other words, the output of the system corresponding to a given input remains the same no matter when the input is fed into the system. Therefore, if there is a time shift of the input of a system, there will also be a time shift in the output side. In (2.3), if $x[n]$ is shifted to the right by n_0 (represented as $x[n - n_0]$), the corresponding output will be $y[n - n_0]$.

A system that is both linear and time invariant is a linear time-invariant system, or an LTI system. The LTI system plays a fundamental role in signal processing because many actual processes are LTI, and such system can be analyzed and characterized with ease with a set of powerful mathematical tools.

2.2.4 Unit Impulse Response and the Convolution Sum

By (2.2), we stated clearly that all discrete time signals are actually a series of scaled impulses. And combining with the linear and time invariant property of section 2.2.3, we imply that if one wishes to know the response of any discrete time signal, one can firstly break the signal down into a summation of impulse signals, then find the responses of each impulse signals, and finally sum these responses to form the full response of the original discrete time signal. In other words, the response produced by any discrete time signal is a superimposed combination of displaced and scaled responses produced by a unit impulse. Let the input be $x[n]$ and let the output be $y[n]$, the input/output relationship of a LTI system can be represented by the following formula:

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]\, h[n-k] \qquad (2.4)$$

Where $h[n]$ is the response of the unit impulse, or simply the "unit impulse response". Equation (2.4) is extremely significant, because it means that if we can have the (unit) impulse response of an LTI system, we can calculate the output of any input signal, even without the presence the actual system itself. That is to say, *an LTI system is fully characterized by its (unit) impulse response.* Equation (2.6) is known as the *convolution sum* of the signal $x[n]$ and the unit impulse response $h[n]$. The convolution operation is represented symbolically as

$$y[n] = x[n] * h[n] \qquad (2.7)$$

The knowledge of LTI systems and the useful nature of the impulse response will serve as the basis of the following chapters. It is further shown in [10] that by discrete time Fourier transform, equation (2.7) can be transformed to the frequency domain as

$$Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega}) \qquad (2.8)$$

where $X(e^{j\omega})$, $H(e^{j\omega})$and $Y(e^{j\omega})$ are the Fourier transforms of $x[n]$, $h[n]$, and $y[n]$.

Chapter 3

Glass Containers – The Scenario and Specifications

3.1 The Scenario

The scenario to be analyzed is set to be the acoustic phenomena in water-filled glass containers with the listener at the nearby outside. This scenario is designed to be tricky in terms of auralisation techniques. Since the containers are much smaller than the buildings and rooms surveyed in section 1.2, traditional ray tracing method and the image source method cannot be utilized. Other facts that are different with the previous studies in room acoustics are that the sound source is placed under water, and the listener is outside of the container. Plus, the glass containers selected is open at the top, as opposed to the closed chambers in room acoustics.

3.2 The Specifications

- **Glass Container 1**

  An ellipsoidal fish tank with –

Physical dimensions (W× L ×H): 22.0 cm × 22.0 cm × 16.5 cm

Glass thickness: 0.2 cm

Diameter of the top opening: 16.5 cm

Diameter of the bottom stand: 11.0 cm

Fig. 3-1 Glass container 1

- **Glass Container 2**

A rectangular fish tank with –

Physical dimensions (W× L ×H): 66.0 cm × 37.0 cm × 36.0 cm

Glass thickness: 0.4cm

Fig. 3-2　Glass Container 2

Chapter 4
Exponential Sinesweep Method

As mentioned in section 2.2, an LTI system is fully characterized by its impulse response. With the assumption that the water-filled glass containers possess near-linear properties, the work to be done for auralisation is therefore to find the impulse response of the containers. The most trivial method to do so is to create an acoustic impulse within the glass container, and record the response with a recorder. However, for obtaining the necessary accuracy and quality of the response, this impulse has to be extremely short in duration and has extremely high pulse energy. These requirements pose problems in the recording end since the capabilities of digital recorders limit them from recording an accurate response. Hence, in this part of the research, we resort to another method developed in the last decade to perform the measurement: the exponential sinesweep method [11].

4.1 The Theory of the Exponential Sinesweep Method

The exponential sinesweep method does not require producing an actual impulse in the glass container; instead, it plays a sinesweep signal. A sinesweep signal is a sinusoidal signal with exponentially varying frequency and constant amplitude. Figure 4-1 is an illustration of an exponential sinesweep signal.

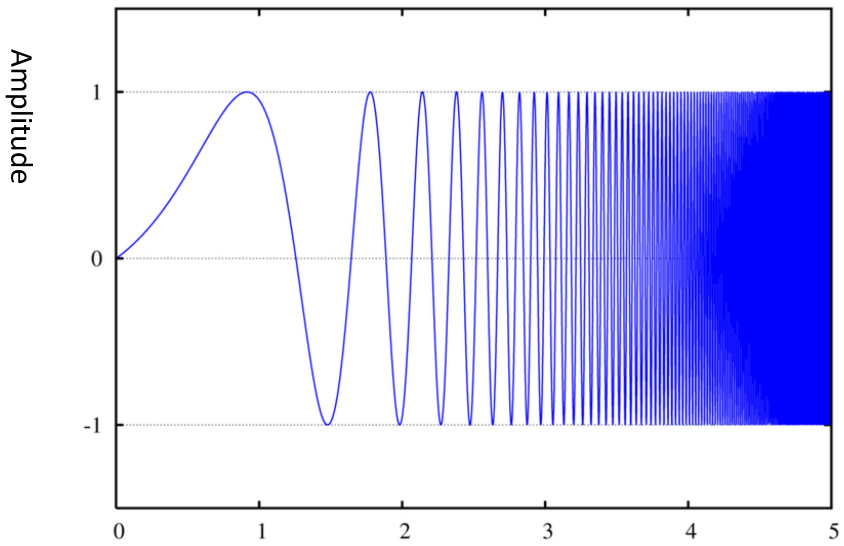


Fig. 4-1 An exponential sinesweep waveform; a sinusoidal wave that increases in frequency exponentially over time. Excerpted from Wikipedia: chirp

The sinesweep signal used in this thesis is similar to the figure above, given mathematically below:

$$x(t) = sin\left[\frac{\omega_1 \cdot T}{ln\left(\frac{\omega_2}{\omega_1}\right)} \cdot \left(e^{\frac{t}{T} \cdot ln\left(\frac{\omega_2}{\omega_1}\right)} - 1\right)\right] \tag{4.1}$$

The signal starts with frequency ω_1 and ends with frequency ω_2. The total duration is T. For the deduction of equation (4.1), the reader is recommended to refer to [11]. In auralisation science, the lower frequency response contributes more to the overall impulse response. As we can see, the exponential sinesweep emphasizes more on the lower frequency range than the higher frequency range, and is therefore more ideal than linear sinesweeps.

This method obtains the impulse response collaterally through a deconvolution technique. Let the input sinsweep signal be $x[n]$, and the output signal be $y[n]$, the impulse response of the system be $h[n]$, the measurement can be represented by

$$y[n] = s[n] + x[n] * h[n] \tag{4.2}$$

Where $s[n]$ is the noise picked up by the recorder during the measurement. The noise in this case in assumed to be completely uncorrelated with the input signal, and is assumed to be white and Gaussian. To avoid the noise from contaminating the desired impulse response, an averaging technique is performed on $y[n]$ prior to deconvolution to increase the signal to noise ratio. Here, we perform the deduction of deconvolution. Let $\hat{y}[n]$ be the output processed by the averaging technique. From equation (2.8), we can deduct that

$$H(e^{j\omega}) = \frac{\hat{Y}(e^{j\omega})}{X(e^{j\omega})} \tag{4.3}$$

And by Fourier transform we get

$$H(e^{j\omega}) = \frac{FFT(\hat{y}[n])}{FFT(x[n])} \tag{4.4}$$

Thus $h[n]$, which is the impulse response of the system, can be obtained by the inverse Fourier transform:

$$h[n] = iFFT\{H(e^{j\omega})\} = iFFT\left\{\frac{FFT(\hat{y}[n])}{FFT(x[n])}\right\} \tag{4.5}$$

4.2 The Experiment Set Up and Procedures

This experiment attempts to acquire the impulse responses of the two glass containers mentioned in Chapter 3.

The hardware equipment used in this experiment is listed as follow:

- Two water-filled fish tanks (see section 3.2 for specifications)
- Sound source: Altec Lansing iM227 compact speaker
- Microphone: Zoom H4n digital recorder, recording in 96 kHz, 16 bit
- Thin kitchen plastic bags
- Tripod: Joby Gorillapod SLR ZOOM
- Alarm Clock

The frequency response of the microphone is given in appendix B for reference.

The procedures of the experiment are presented below:

1. Fill the glass tank 80% full with room temperature water.
2. Seal the compact speaker tightly with a thin plastic bag.
3. Connect the speaker with a computer used to play the sinesweep.
4. Place the speaker in to the container; submerge it into the water.
5. Mount the microphone onto the tripod, facing the speaker output.
6. Play and record the sinesweep
7. Repeat step 4 to step 6 with the empty container.
8. Preform deconvolution on the recorded sinsweep to acquire the impulse response.
9. Perform the auralisation by convolving the sinesweep with desired sounds.

Figure 4-2 and Figure 4-3 are the pictures of water filled glass tanks with the sound source

Fig. 4-2 Glass container 1 with sound source

Fig. 4-3 Glass container 2 with sound source

Note that the recording environment has to be as quiet as possible to avoid noise contamination.

The original sound files used for auralisation are also played directly in the water-filled tanks and recorded for comparison. The sound files include the recordings of a ringing alarm clock and human spoken words. In addition, an actual ringing alarm clock is also placed in the tank, and recorded for comparison,

Chapter 5
Finite Element Method

5.1 Finite Element Method and COMSOL Multiphysics

A brief introduction to finite element method (FEM) was already given in section 1.2. FEM is a numerical method that subdivides an object, either 2D or 3D, into small but finite pieces. When a simulation scenario takes place, the corresponding physical and mathematical equations governing the behavior of the system is applied to each and every small piece simultaneously. The calculated results of these pieces are summed up to give a macro description of how the system will behave. One of the reasons of attempting to utilize FEM as a means of auralisation is that no physical object has to be built; everything is done with simulations on 2D or 3D models via computer software. This may be of great advantage if the scenario to be auralised is hard to construct. Furthermore, in the workflow of computer animations, 3D models are often readily available.

In this research, a numerical simulation software, COMSOL Multiphysics, is utilized. COMSOL Multiphysics is a tool for modeling and simulating various physical problems base on FEM and partial differential equations (PDEs). The user of COMSOL Multiphysics does not need to have a complete knowledge of FEM or the underlying PDEs. Instead, the user takes the ease of just performing the following steps: determine the type of the physics problem; construct the 2D or 3D model of the problem; describe the physical properties, the boundary conditions, and the imposed forces of the problem. The main problem of COMSOL Multiphysics is that sometimes when the simulation requires highly accurate results, the subdivision of the model has to be considerably complex. That is to say, it may sometimes require a large amount of subdivided pieces to perform the simulation. This may result in an extremely computationally intensive process.

5.2 Modeling Procedures

Similar to Chapter 4, the goal of the FEM method is to find the impulse response of the scenario. Here we model the water-filled 3D glass container with a point sound source submerged. In this case, the source produced a Gaussian sound pulse.

The hardware and software utilized is listed below:

- Hardware: Personal computer running Windows 7 64 bit with Intel core i5-750 quad-core CPU and 16 GB of DDR3 memory.
- Software: COMSOL Multiphysics 4.2a

The modeling procedures are described below. Tedious details and software settings are omitted.

1. Specify the problem type as **"Acoustic-Solid Interaction, Transient"**
2. Construct the 3D model of the scenario based on the actual dimensions of the glass container.
3. Determine the position of the sound source.
4. Define the outer boundary of the problem. In this case the glass container is placed in an air-filled large sphere.
5. Specify the materials of the model.
6. Specify the boundary conditions and determine whether the domains propagate acoustic wave (in fluids) or elastic wave (in glass).
7. Specify the intensity and center frequency of the Guassian pulse.
8. Mesh the scenario in tetrahedral. Make sure that each wave is resolved.
9. Compute the problem

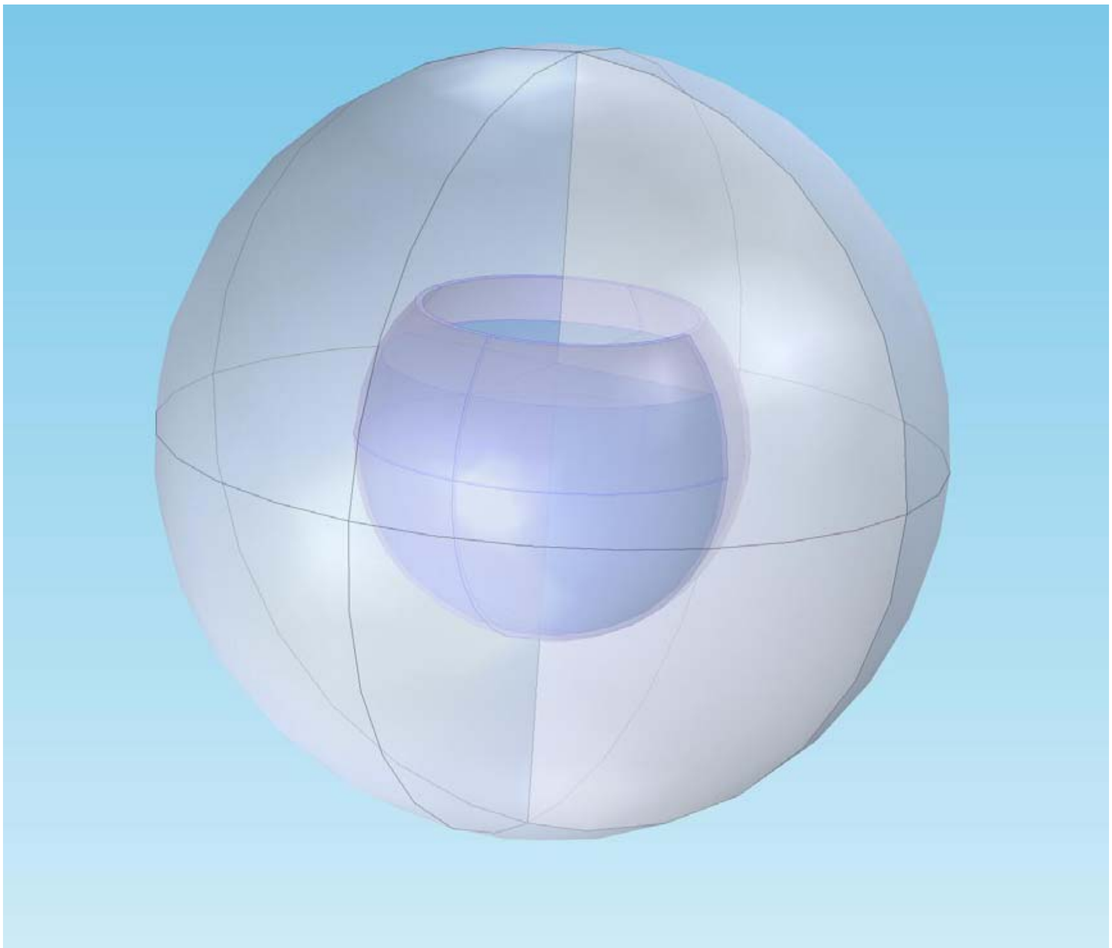


Fig. 5-1 3D model of Glass container 1 in COMSOL; the outer sphere is an air domain.

Chapter 6

Results and Discussions

6.1 Recordings of Sound Played in Glass Containers

Several sound clips and a ringing alarm clock are used as sound sources in glass container 1. The resulting sound seems suppressed in the higher frequency domain. The case of the actual alarm clock is the most obvious: the high frequency sound seems drowned, and the low frequency sound is therefore more audible. This happens in human voice sound sources too. Furthermore, in all cases, the sound volume decreases dramatically, as expected in [14].

6.2 The Results and Discussions from the Exponential Sinesweep Method

All the results mentioned in this section can be found and heard via soundcloud.com:

http://soundcloud.com/auralisation_jcmg

The impulse response of glass container 1 is acquired in two cases: with and without water in the container.

- **Impulse response without water**
- **Impulse response with water filled**

The frequency response of the wet tank is shown below. As predicted in section 6.1, the water-filled tank 1 has the characteristics of a low pass filter.

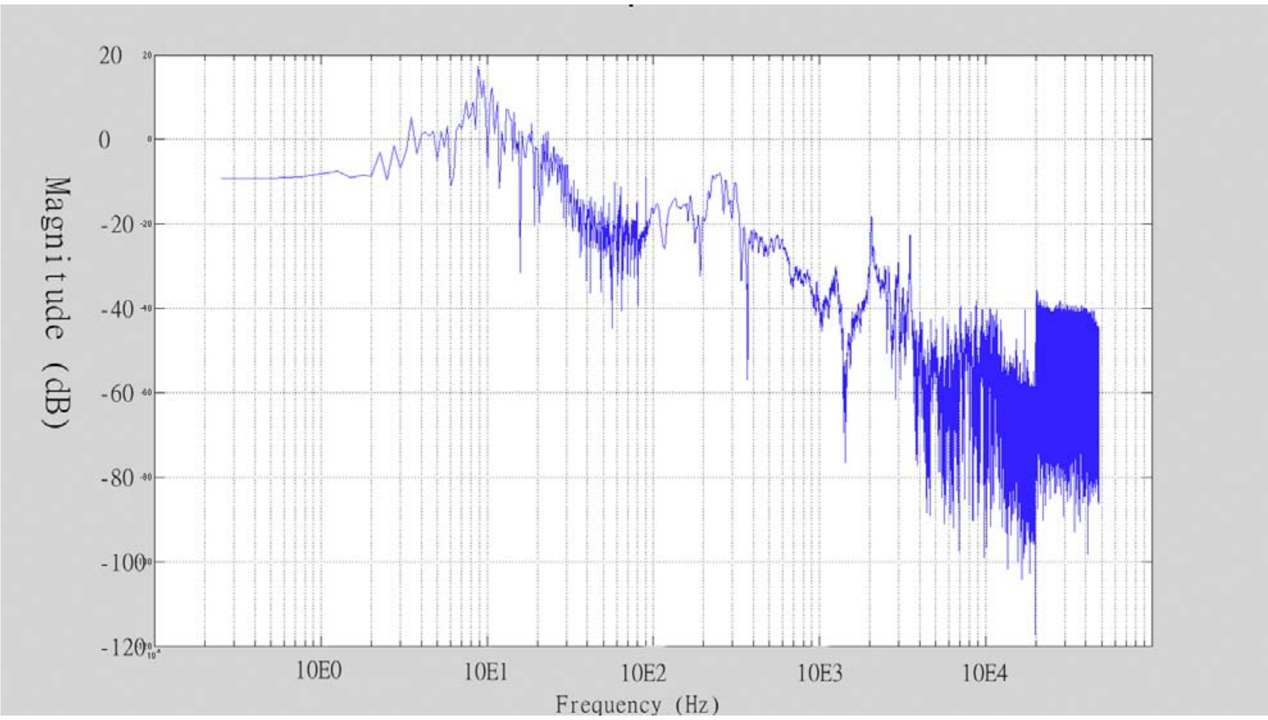


Fig. 6-1 Frequency response of glass container 1

The transient response plot of the water-filled tank 1 is shown in Fig. 6-2:

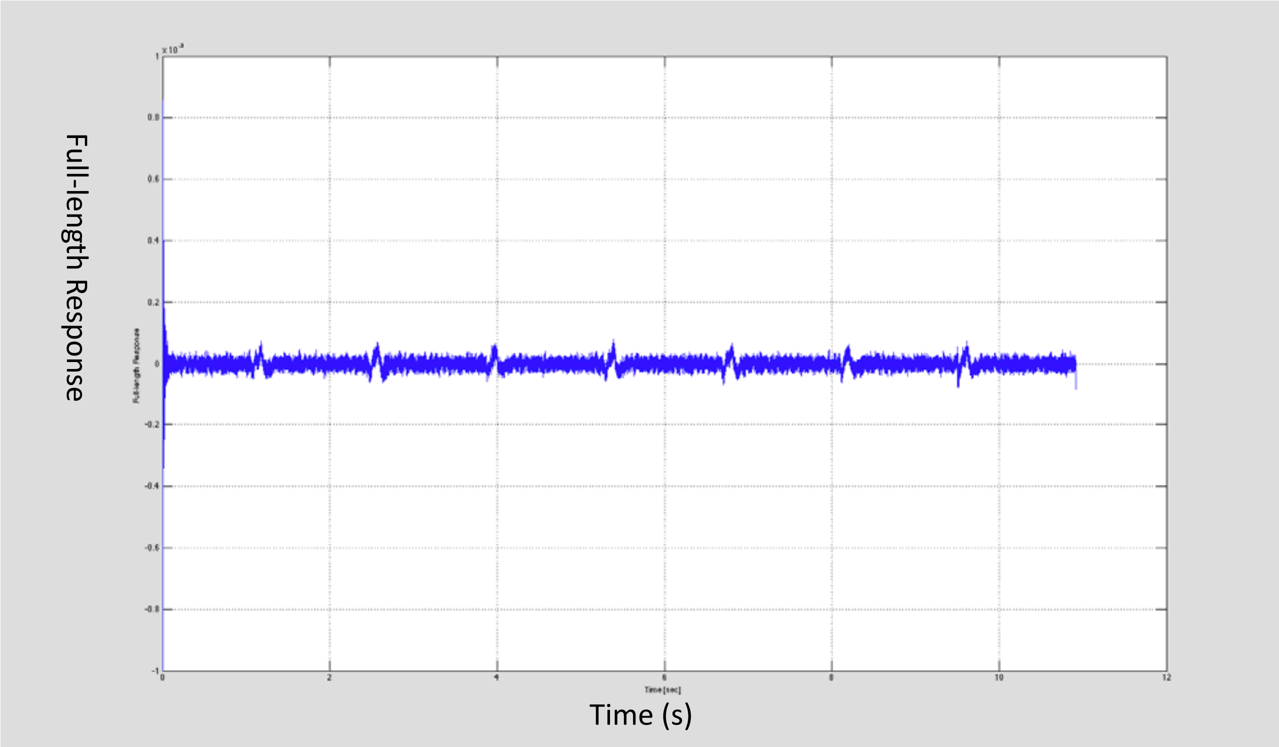


Fig. 6-2　Transient response of glass container 1

The two original sound clip to be auralised can be heard as well:

- **Human voice original sound clip**
- **Alarm clock original sound clip**

Convolving the above two sound clips with the impulse response of water-filled tank, the auralised sound is presented:

- **Auralised human voice in water-filled glass tank**
- **Auralised alarm clock sound in water-filled glass tank**

For reference, these are the actual recordings of these sound files played by a submerged speaker:

- **Real recording of human voice played in water-filled glass tank**
- **Real recording of alarm clock sound played in water-filled glass tank**

We can hear that the auralisations of these scenarios are quite successful in these cases; the recording and the auralisation quality are directly related to the quality of the recording studio. Notice that the plastic bag is utilized both in the making of the impulse response, and the direct recording of the submerged speaker. In other words, the two cases are physically equivalent. For comparison, the auralised sounds for the dry tank case are presented below:

- **Auralised human voice in dry glass tank**
- **Auralised alarm clock sound in dry glass tank**

To produce a real case application in computer animation, a clip from Pixar animation *Finding Nemo* (2003) [12] is selected. The scene involves the case of fishes inside a glass tank talking with a Gull outside. In the original sound track, there are no differences in sound treatment between the fishes' voices in the tank and the Gull's voice. To utilize the result of this research, the soundtrack is edited by replacing the fishes' voices with their auralised counterparts. The impulse response of the water-filled tank (glass tank 1) is used in this auralisation. The result is presened at: http://www.youtube.com/watch?v=3caw1DuLiUI

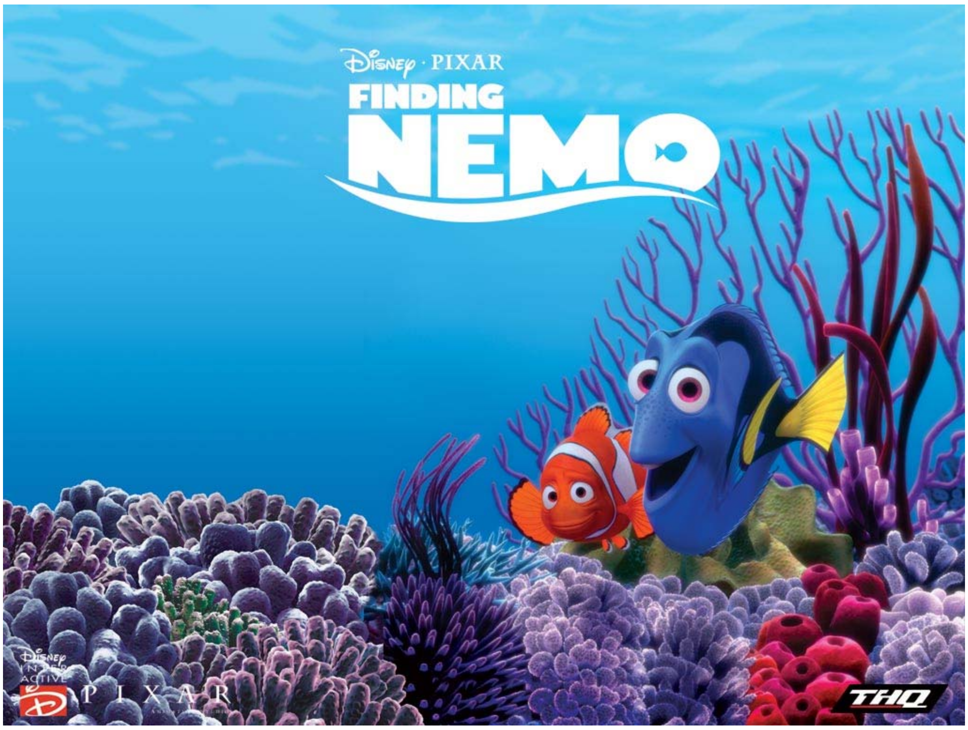


Fig. 6-3　Poster of the 2003 Pixar animation *Finding Nemo*

As we can see, this treatment in sound makes the dialog more understandable. Notice that the shape can size of glass tank one is very different from the fish tank in Finding Nemo, but perceptually, the sound works fine in the animation.

For the larger glass tank 2, however, the impulse response is not acquired successfully, and can be heard from soundcloud.com.

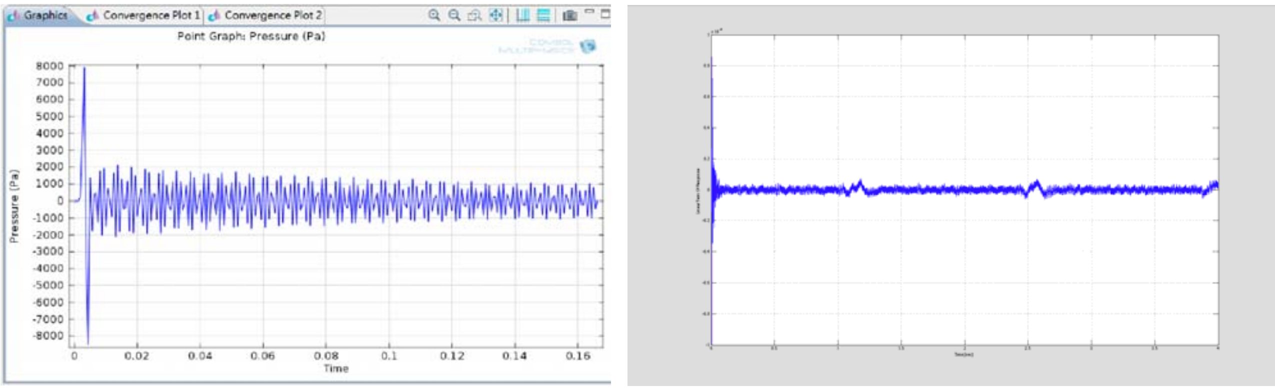
Although the exponential sinesweep method is designed to tolerate weakly non-linear systems, the experiment has shown that it does have its limits. The non-linearity of the frequency response of glass tank 2 can be observed at the lower frequencies where sound with frequencies higher than the input is audible in the output. This research chose a basic form of the exponential sinesweep method. Over the past decade, there were improvements made on the method for dealing with non-linearity and noises. It may be possible to make this method more robust if improved algorithms are adopted.

Nevertheless, for real world applications in the creative industries, this basic exponential sinesweep method still offers lots of opportunities. The perception is the most important factor in sound designing, not accuracy. With this method, the sound designers' horizon of possibilities is widened, as a new group of sound effect may be created through impulse responses.

6.3 The Results and Discussions from the Finite Element Method

The FEM method did not produce satisfactory impulse response results. Since the thickness of the walls of the glass containers are very thin comparing with their physical dimensions, meshing the thin layer of glass requires very small tetrahedral to produce decent results. However, as the size of these tetrahedral decreases, the computational intensity becomes impractically high. Since the thickness of the tanks has great effects on their frequency responses, the simulation will have little meaning if one takes compromise by modeling a thicker tank.

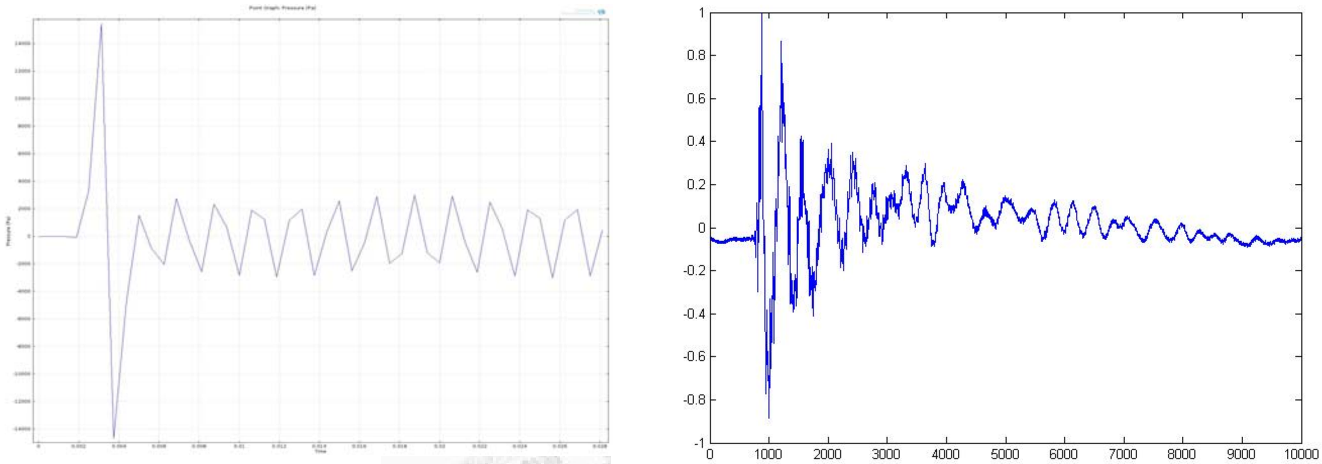
Nevertheless, in the case of glass tank 1, it we choose to compromise and model the tank with a wall thickness of 2 cm (as opposed to the real case of 0.2 cm), we can still get a seemingly similar response comparing with the result of the exponential sinesweep method.



(a) (b)

Fig. 6-4 Comparison between (a), Transient response of glass tank 1 generated by FEM (a), and exponential sinesweep method (b)

However, as we focus on the beginning of the transient responses, the most crucial part of the response, we can see that they behave very differently:



(a) (b)

Fig. 6-5 Details of the initial stages of the responses generated by FEM (a), and exponential sinesweep method (b)

It is worth noticing that in this experiment, even with one compromised with the wall thickness and set it to 2 cm, it still takes several days to calculate the response.

For comparison, like glass tank 1, the Finding Nemo sound track is also rendered with this response: http://youtu.be/NP7uKcCza78

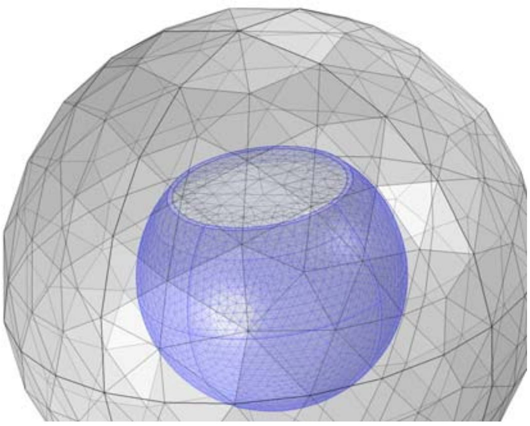


Fig. 6-6 The FEM mesh of the 0.2 cm container 1

Chapter 7
Conclusions

The possibilities of applying modern science techniques to sound design in the creative field has been explored and demonstrated. In this thesis, attempts have been made to perform auralisation on a specific case – sound sources in water-filled glass containers, by the exponential sinesweep method and the finite element method. A demonstration on a modern animated film has been made based on the results produced by the exponential sinesweep method. Although the exponential sinesweep method may produce very decent and accurate auralisation results, this study has shown that it does have its limits. However, such limits may be leveraged by means of improving the algorithm base on recent studies. Concerning the finite element method, the computational intensity required is still impractical in this case, yet it might be proven useful in other cases which do not require a dense mesh. As the computational power of modern computers improves, there should be chances that this method would open up a new world of possibilities, given the fact that it can create and simulate scenarios that can only be imagined. With the aid of auralisation technology, it is possible for sound designers to build an entire library of impulse responses. These impulse responses are reusable and can be powerful tools when there is a need to create realistic soundtracks. Finally, one should bear in mind that the application of these methods should not be limited to a certain field, such as sound designing in animations. Other acoustic related fields such as music composition may utilize the unique sounds that science can offer, and take our listening experiences to a brand new level.

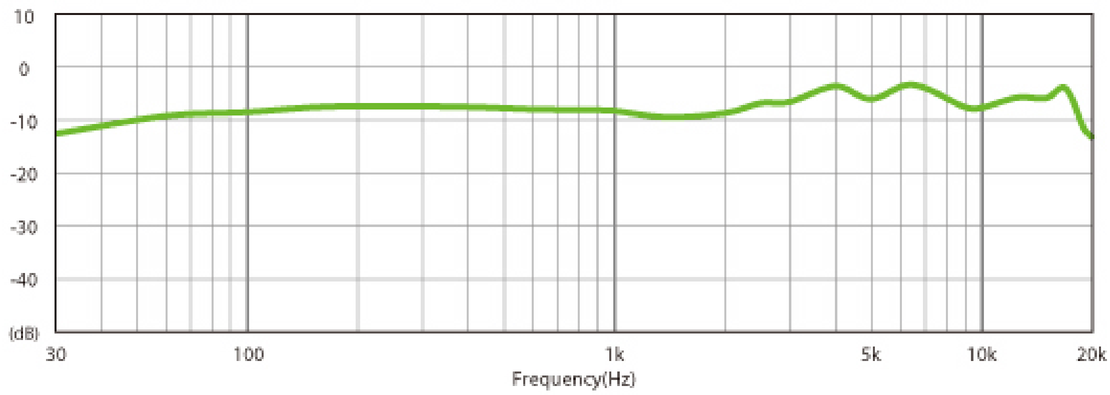
Appendix

- **Appendix A**

As suggested by [1], the definition of "auralisation" is:
"Auralisation is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space."

- **Appendix B**

Below is the frequency response of ZOOM h4N microphone provided by the manufacturer:



Bibliography

[1] Kleiner, M., B. I. Dalenback, et al. (1993). "Auralization - an Overview." Journal of the Audio Engineering Society **41**(11): 861-875.

[2] Xiang, N. and J. Blauert (1993). "Binaural Scale Modeling for Auralisation and Prediction of Acoustics in Auditoriums." Applied Acoustics **38**(2-4): 267-290

[3] Oguchi, K., S. Ikeda, et al. (1993). "Application of Binaural Hearing to Scale-Model Testing." Journal of the Audio Engineering Society **41**(11): 931-938.

[4] Easwaran, V. and A. Craggs (1995). "On Further Validation and Use of the Finite-Element Method to Room Acoustics." Journal of Sound and Vibration **187**(2): 195-212.

[5] Craggs, A. (1986). "A Finite-Element Model for Acoustically Lined Small Rooms." Journal of Sound and Vibration **108**(2): 327-337.

[6] S. Sakamoto, H. Nagatomo, A. Ushiyama and H. Tachibana (2008). "Calculation of impulse responses and acoustic parameters in a hall by the finite-difference time-domain method,'' Acoust. Sci. & Tech., **29**, 256–265.

[7] Bonfiglio, P. and Farnetani, A (2006). "Low frequency analysis of small rooms by means of a finite element model", Proceedings of the COMSOL Users Conference 2006 Milano

[8] Milson, E., B. Burtt et al. (2008) Animation Sound Design: Building Worlds From The Sound Up, DVD featurette, Walt Disney Studios Home Entertainment

[9] Jackson, P., B. M. Osborne, et al (2002). The Lord of the rings--The Two Towers, motion pictures DVD featurette, New Line Cinema.

[10] Oppenheim, A. V., A. S. Willsky, et al. (1997). Signals & systems. Upper Saddle River, N.J., Prentice Hall.

[11] Angelo Farina (2000). "Simultaneous Measurement of Impulse Response and

Distortion with a Swept-sine technique", *J.AES*, vol. 48, p. 350, 108th AES Convention, Paris, Preprint 5093

[12] Newman, T. and R. Williams (2003). Finding Nemo. Burbank, CA, Walt Disney Records

[13] Kinsler, L. E. (1982). Fundamentals of acoustics. New York, Wiley

[14] Dowling, A. P. and J. E. Efowcs Williams (1983). Sound and sources of sound. Chichester, New York, E. Horwood ; Halsted Press.