2007-03-21

# Objective Test Methods for Waveguide Audio Synthesis

Steven Gregory Wood
*Brigham Young University - Provo*

OBJECTIVE TEST METHODS FOR WAVEGUIDE AUDIO SYNTHESIS

by

Steve G. Wood

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

School of Technology

Brigham Young University

April 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Steve G. Wood

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

| | |
|---|---|
| Date | Michael G. Bailey, Chair |
| Date | Ronald F. Gonzales |
| Date | Barry M. Lunt |

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Steve G. Wood in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Michael G. Bailey
Chair, Graduate Committee

Accepted for the School

Barry M. Lunt
Graduate Coordinator

Accepted for the College

Alan R. Parkinson
Dean, Ira A. Fulton College of Engineering and Technology

ABSTRACT

OBJECTIVE TEST METHODS FOR WAVEGUIDE AUDIO SYNTHESIS

Steve G. Wood

School of Technology

Master of Science

Acoustic Physical Modeling has emerged as a newer musical synthesis technique. The most common form of physical modeling synthesis in both industry and academia is digital waveguide synthesis. Commercially available for the past thirteen years, the top synthesizer manufacturers have chosen to include physical modeling synthesis in their top of the line models.

In the area of audio quality testing, the most common tests have traditionally been group listening tests. While these tests are subjective and can be expensive and time-consuming, the results are validated by the groups' proper quality standards. Research has been conducted to evaluate objective testing procedures in order to find alternative methods for testing audio quality. This research has resulted in various standards approved by the International Telecommunication Union. Tests have proven the

reliability of these objective test methods in the areas of telephony as well as various codecs, including MP3.

The objective of this research is to determine whether objective test measurements can be used reliably in the area of acoustic physical modeling synthesis, specifically digital waveguide synthesis. Both the Perceptual Audio Quality Measure (PAQM) and Noise-To-Mask Ratio (NMR) objective tests will be performed on the Karplus-Strong algorithm form of Digital Waveguide synthesis. A corresponding listening test based on the Mean Opinion Score (MOS) will also be conducted, and the results from the objective and subjective tests will be compared.

The results will show that more research and work needs to be done in this area, as neither the PAQM nor NMR algorithms sufficiently matched the output of the subjective listening tests. Recommendations will be made for future work.

# ACKNOWLEDGMENTS

I am extremely grateful for all of the help, support, and encouragement that everyone gave me during this thesis. Thank you to my committee for the guidance, advice, and lessons. Especially thanks to my chair Dr. Bailey for taking so much time to meet with me, and for the constant encouragement.

There were many people in industry and academia that took time to answer various questions that I had. Thank you to all of them, especially Jeremy Geisler from Digitech.

Thanks to my family, in particular my brother Jeff for helping me with the samples I needed.

Most of all, I am grateful for my wife Crissy and son Parker who were extremely encouraging and patient while I took on this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# 1 Introduction

## 1.1 Background

Generic instrument sound replication technologies have been used since the nineteen-eighties. These generic sounds have used various synthesizer technologies, including:

- Analog Synthesis
- Frequency Modulation
- Sampling
- Additive Synthesis
- Physical Modeling

These technologies, which also include systems combining different synthesizer technologies, have had the ability to create the sounds of various instruments, including pianos, guitars, woodwinds, brass, percussion, etc. Only recently has the technology advanced to include specific instruments as opposed to generic ones. A few products, including the Roland GS system, the Digitech GNX processors, and the Line 6 POD units, have been introduced with the ability to recreate specific sounds and characteristics of selected instruments, including particular guitars and amplifiers, rather than broader categories of instruments.

Acoustic physical modeling measures use the processes associated with a complex computer algorithm that replicates and/or synthesizes the sonic and performance behaviors of an actual acoustic instrument (Tapia, pg. 1). Acoustic modeling of musical instruments is a technology that has existed for over two decades. The technology of replicating modeling between instruments has advanced significantly during the past thirteen years. This has been made possible since the introduction of the Yamaha VL1 in 1994 (Physical Modeling Synthesis, 2005). The VL1 was the first commercially available synthesizer to use physical modeling technology. The VL1 used physical modeling to model pipe and string resonances (B. Angelos, personal communication, February 6, 2006). There are now several manufacturers producing synthesizers that use physical modeling technology to model brass, woodwinds, percussion, strings, pianos, and guitars, among others.

Physical modeling synthesis includes various techniques used to model instruments, such as waveguide synthesis, finite element methods, modal synthesis methods, and banded waveguides, with waveguide synthesis being the most popular technique so far (Rabenstein, Trautmann, 2002, pg. 3). Each of these physical modeling technologies has their own strengths and weaknesses. For example, finite element methods use fewer delays than waveguides, making them quicker to run with fewer computations, and therefore, more economically efficient. However, due to their dual-delay feedback finite elements methods are also less stable, as the feedback loop has the potential to infinitely grow (Karjalainen, 2004, pg. 11).

Of significant interest in musical instrument modeling is the ability to determine which technique will provide the optimal synthesis for a specific instrument. Identifying a quantitative objective measurement standard for the purpose of comparing modeling methods associated with musical instrument sound qualities would enhance the profession as newer concepts of musical instrument modeling are developed and implemented. Included in this research question of determining optimal replication is the method of measurement and assessment used to identify the perceptual quality of the physical modeling synthesis-based sound simulations. This area of research was recently discussed in Georg Essl's 2002 doctoral dissertation entitled "Physical Wave Propagation Modeling for Real-Time Synthesis of Natural Sounds." This dissertation mentioned measurement processes using complex mathematical formulas measuring octaves, wavelengths, amplitudes, distortion, wow, and flutter. These mathematical terms have been applied to properties associated with musical tone variations subjectively arrived at through group listening tests. Identifying an objective measurement standard for the purpose of evaluating perceptual audio quality and replacing subjective testing would be beneficial in the areas of both industry and academia.

## 1.2 Problem Statement

The purpose of this research is to test objective audio quality measurement standards used in other fields against audio physical modeling synthesis. The results from these tests will be compared to results from a subjective listening test, in order to validate the objective audio quality tests.

## 1.3 Hypothesis

By comparing measured objective results to those subjective results from a qualified panel of listeners, this research will determine if either the Perceptual Audio Quality Measure (PAQM) system and/or the Noise-to-Mask Ratio (NMR) are valid objective measurement techniques for quality of audio synthesized by physical modeling techniques, specifically waveguide synthesis. If the total results from the objective tests fall within ninety percent accuracy of the listening test results, the objective tests will be defined as being valid. This thesis will follow an exploratory research design, specifically using predictive research in an attempt to find relationships between the objective test algorithms (PAQM, and NMR) and the subjective listening tests. It is expected that both the PAQM and NMR systems will prove to be valid metric measurement systems when used with the physical audio modeling synthesis technique of waveguide synthesis. Both the PAQM and NMR systems have already proved useful as measurement systems in the areas of digital speech coding in wireless telephony, as well as in music coding for compression (Essl, 2002, pg. 128).

## 1.4 Justification

There is a need to identify a metric or measurement system to determine the perceptual quality and accuracy of physical modeling techniques (Van den Doel, Pai, 1996, pg. 5). Jeremy Geisler of Digitech (2005) stated that ninety percent of the audio testing done in the music industry is recognized audibly. There have been studies as to which metrics take the least computations, which translates into economics, but the most important factor in musical audio is the final sound (J. Geisler, personal communication, June 22, 2005). Some research has been conducted to study perceptual characteristics of physical modeling techniques (Jarvelainen, 2001, pg. 1). However, with regards to perceptual quality testing of physical modeling techniques, Georg Essl (2005) said "…there are many openings that are largely unexplored." This research will provide project managers, engineers, researchers, developers, and other interested groups with information and metrics used to compare how the PAQM and NMR systems perform with physically modeled audio.

### 1.5 Methodology

In this research, sounds will be synthesized using the Max/MSP program from Cycling 74, controlled with a MIDI controller. Each sound will be captured using Cubase SX 2.0 software. Original instrument sounds will also be captured using Cubase in order to perform comparisons between original and synthesized sounds. These comparisons will be carried out using the PAQM and NMR algorithms. Both the PAQM and NMR algorithms require Digital Signal Processing operations, such as FFT's, frequency to pitch transformations, and frequency and time domain smearing/spreading; metrics associated with tone validation. These operations will be performed using Matlab software. A measurement listening test will also be performed in order to form comparisons, to validate and evaluate the results from the PAQM and NMR tests. The listening panel will be made up of individuals currently or previously enrolled in music classes, or with other ties to the music industry. Statistics will be performed to show the repeatability, and reliability of the results. If the results from the PAQM and NMR test fall within one standard deviation of the results from the listening test, they will be defined as being valid, objective test measurements.

## 1.6 Assumptions

When musical instruments are recorded or sampled, there is a certain amount of noise introduced. Recording direct without any microphones will help minimize the introduction of noise, but ultimately the noise cannot be completely eliminated. However, the noise is inaudible, and therefore it is assumed that the noise introduced into the original samples is negligible, and will not have an effect on the overall analysis.

## 1.7 Delimitations

To complete the necessary audio processing and comparisons, it is necessary to limit the type of audio sounds and synthesis techniques used in the study. Therefore, plucked strings will be the only sounds that are used in this study, and waveguide synthesis will be the exclusive physical modeling technique tested. Waveguide synthesis is the most common, and the most important physical modeling technique used in the industry and academia (Hiipakka, 1999, pg. 2).

# 2 Review of Literature

## 2.1 Physical Modeling Techniques

Physical modeling synthesis is a way of creating audio using a numerical model of a physical system. This is different from other types of musical synthesis because physical modeling algorithms attempt to represent the characteristics of specific instruments using a model based on a physical analysis of the system. Other synthesis methods use filters and waveform generators to try and imitate the timbre of a real instrument. Physical modeling synthesis generates sound that refers to the actual physical response of the system (Ashman, 2002). When first introduced, the sound quality of physical modeling synthesis was considered to be quite inferior to sampling technology (Karjalainen, Valimaki, Janosy, 1993, pg. 1). However, it is now considered to be one of the better synthesis techniques available, and is usually only included on professional level synthesizers.

There are many different physical modeling techniques. These methods can be divided into four groups: waveguide synthesis, finite element methods, modal synthesis methods, and banded waveguides. Other synthesis techniques that don't fit in these categories are either stochastic in nature, such as dynamic stochastic synthesis, or are not physically informed, such as sampling (Roads, 1996, pp. 340-342). In practice, researchers usually have their favorite acoustic synthesis methods. However, occasionally different methods are combined and used together (Karjalainen, 2004, pg. 9).

### 2.1.1 Waveguide Synthesis

Waveguide synthesis focuses on the resonant behavior and response from instruments. Waveguide synthesis methods frequently are based on a traveling wave, usually modeled using delay lines (Essl, 2002, pg. 8). Some examples of waveguide synthesis techniques include Karplus-Strong Synthesis, McIntyre, Schumacher, and Woodhouse Synthesis, and Digital Waveguide Synthesis.

#### 2.1.1.1 Digital Waveguide Synthesis

Digital Waveguide Synthesis is a time-domain modeling technique based on the simulation of acoustic wave propagation through a medium using bi-directional delay lines. Each delay line is designated for a directional wave component (Karjalainen, 2004, pg. 2). Dispersion and frequency dependant losses are represented by digital filters.

A lossless digital waveguide takes on the discrete form of d'Alembert's traveling wave solution of the wave equation (one dimensional), as the superposition of waves traveling right and left,

$$y(m,n) = y^{+}(m-n) + y^{-}(m+n), \tag{2-1}$$

where generally $y^{+}$ represents traveling waves going right, and $y^{-}$ represents traveling waves going left. Observing function $y(m,n)$ at point m and time n, one will perceive two delayed copies of the traveling waves (Smith, 2006, Traveling-Wave Solution).

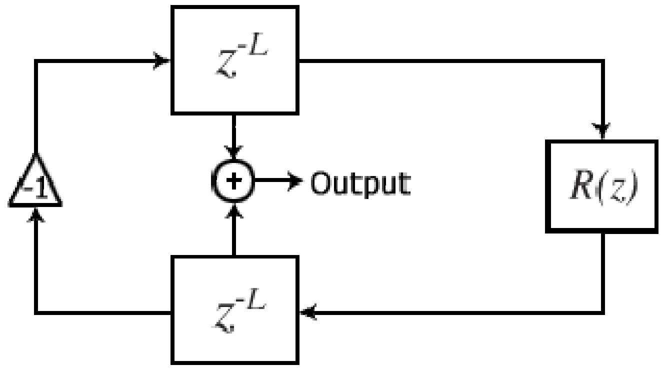


**Figure 2-1: A basic one-dimensional waveguide (likely of a string) with a rigid termination on one end (left) and a frequency-dependent attenuating filter at the other (right). The $z^{-L}$ represents an $L$-sample delay. (Digital Waveguide Synthesis, 2006)**

Figure 2-2 represents the simulation of a lossless, digital waveguide with observation points at x=0 and x=3X=3cT. The symbol $z^{-1}$ represents a one-sample delay. The upper portion represents the delay line of component $y^{+}$ going right, with its input $y^{+}(n)$ on the left, and its output $y^{+}(n-m)$ on the right. The lower portion represents the delay line of component $y^{-}$ with its input $y^{-}(n+m)$ on the right, and its output $y^{-}(n)$ on the left (Smith, 2006, Digital Waveguide Model). $y(nT,0)$ is the input excitation (such as a pluck of a string) and $y(nT,3X)$ is the output (such as string velocity).

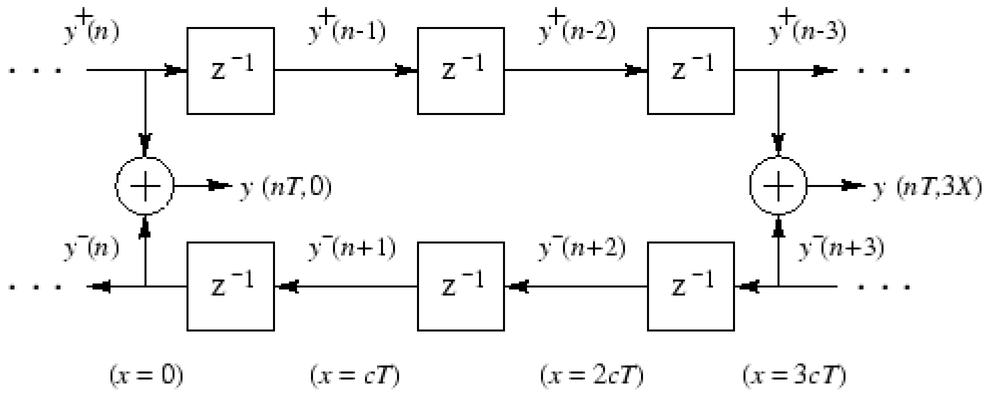


**Figure 2-2: Waveguide digital simulation (Smith, 2006, Digital Waveguide Model), and (Smith, 1992, pg. 4)**

A digital waveguide model of a string comes from the discretization of the traveling-wave solution of the wave equation, as mentioned above. By assuming linearity, all the dispersion and losses of the string with the terminations can be grouped into one point in the model. This way the model can be reduced to a delay line and a filter in a feedback loop (Bank, Valimaki, 2003, pg. 1).

In an earlier paper titled Plucked-String Synthesis Algorithms with Tension Modulation Nonlinearity (1999, pp. 977-980), Karjalainen investigated nonlinear vibrating strings using digital waveguide modeling techniques, where the nonlinearity is caused by tension modulation (Karjalainen, 1999). Karjalainen derived "synthesis models where the nonlinearity is implemented with a time-varying fractional delay filter" (Karjalainen, 1999, pg. 977). This resulted in realistic synthetic tones with nonlinear effects through introducing minor amendments to a linear string synthesis algorithm.

Digital Waveguide synthesis isn't as mathematically obfuscated as some other techniques, but it does take more calculations. This makes it slower than other methods. However, it is also considered to be one of the more stable methods when compared to other physical modeling methods, due to the fact it only uses a single feedback loop

(Karjalainen, 2004, pg. 10). It is also considered to be one of the most efficient physical modeling synthesis methods (Erkut, Valimaki, 2000, pg. 769), and (Laurson, Erkut, Valimaki, Kuushankare, 2001, pg. 38). Digital Waveguide Synthesis is commonly used to model strings, bores, horns, plates, and acoustic spaces (Smith, 1996, pg. 45).

#### 2.1.1.2 Karplus-Strong Synthesis

Stringed instruments (especially guitars) are some of the most simulated and synthesized musical instruments. One of the ways that strings can be simulated by physical modeling is by way of the Karplus-Strong Synthesis (Steiglitz, 1996, pg. 107).

Karplus-Strong Synthesis is based on feedback loops, such as comb filters. Waveforms are looped through a filtered delay line, which simulates the sounds of plucked or hammered strings. Thus, the Karplus-Strong synthesis is primarily used with string sounds, along with certain types of percussion.

The Karplus-Strong method starts out with the generation of a short excitation waveform. The excitation serves as the output, but at the same time is fed back into a delay line. The delay line is typically the same sample length as the excitation waveform. The delay line output then passes through a filter, and then simultaneously goes back into the delay line and back into the output (Yerrick, Karplus-Strong string synthesis, 2006).

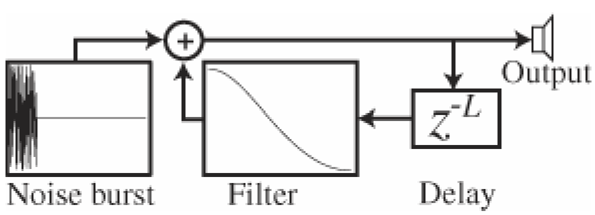


**Figure 2-3: Karplus-Strong Diagram (Yerrick, Karplus-Strong string synthesis, 2006)**

In their book *Music and Computers: A Theoretical and Historical Approach* (2004, pg. 4.9), Phil Burk, Larry Polansky, Douglas Repetto, Mary Roberts, and Dan Rockmore discuss an additive synthesis model (most likely a variation of the reassigned and bandwidth-enhanced modeling technique discussed by Fitz and Haken in "On the Use of Time-Frequency Reassignment in Additive Sound Modeling", 2002) that has been used since the 1980's for modeling plucked strings. This modeling technique uses the Karplus-Strong algorithm. The authors explain how this technique works:

> "…first we start with a buffer full of random values. Noise.…The numbers in this buffer represent the initial energy that is transferred to the string by the pluck. To generate a waveform, we start reading through the buffer, and using the values in it as sample values. If we were to just keep reading through the buffer over and over again, what we'd get would be a complex, pitched waveform.…The pitch we get is directly related to the size of the buffer we're using, since each time through the buffer represents one complete cycle of the signal" (Burk, et al., 2004, pg. 4.9).

The key to the Karplus-Strong algorithm is that every time a value is read from the buffer, it is averaged with the last value that was read. This averaged value is used as the output sample, and the sample is fed back into the buffer. Over time the buffer gets inversely averaged after each sample. Also when the values are averaged, this acts as a natural, first-order low-pass filter on the signal, which in effect limits the number of high frequencies that it contains. This gives the sample a more realistic sound of a real string being plucked. Also as the signal is continually averaged, eventually it will "average out" completely, resulting in a flat waveform. This signals that the string has died out (Burk, Polansky, Repetto, Roberts, Rockmore, 2004, pg. 4.9). The end result is "an initially complex, periodic waveform that gradually becomes less complex over time and ultimately fades away" just like a real string. The periodicity comes from the sample length of the delay line loop, which is represented by the buffer. Burk, et al., do an excellent work describing in detail the Kurplus-Strong algorithm, and how it accurately models an element of a string instrument.

Karplus-Strong Synthesis has existed since the early 1980's and has been primarily used to simulate plucked strings. It was very popular because it was easy to implement, and it didn't require much processing power (Masri, 1996, pg. 20). However, researchers eventually generalized and refined the Karplus-Strong algorithm and came up with the Digital Waveguide Synthesis technique described above (Karjalainen, Valimaki, Tolonen, 2119, pp 1-15). Digital Waveguide Synthesis is considered to be more efficient, and also extends the ability into modeling acoustic waves in tubes and on drums (Smith, 2006, History of Enabling Ideas), but Karplus-Strong Synthesis is still very prevalent in both industry and academia.

#### 2.1.1.3 McIntyre, Schumacher, and Woodhouse Synthesis

McIntyre, Schumacher, and Woodhouse stated that many instruments can be described as linear resonators, modeled by waveguides or all-pole resonators, and a single nonlinear oscillator (for excitation). The separation of the linear resonator and the nonlinear excitation in a time-domain view provides a simulation that models the tonal changes in dynamics very well (Essl, 2002, pg. 9). Examples of some nonlinear oscillators would be a clarinet reed, a flute jet, the bow-to-string friction on a violin, and a brass player's lips (Cook, 2002, pg. 122). These nonlinear elements provide excitation in the form of impulse waves which are then sent to the linear, resonant part of the instrument being modeled. The linear portion then acts as a filter to shape the waveform into the timbre characteristic of the instrument (Roads, 1996, pg. 281).

McIntyre, Schumacher, and Woodhouse Synthesis is considered to be an efficient, simple synthesis technique having an advantage that the parameters for control are closely related to those used by the performers. However, due to its simplifications, this technique is not as realistic as some others (Roads, 1996, pp. 279-281).

### 2.1.2 Finite Element Methods

Finite Element Methods are another type of physical modeling technique that focuses on direct discrete simulation of local dynamics responsible for sound generation (Essl, 2002, pp. 9-10). In the area of music synthesis, finite element methods are used in two ways. First is the study of musical acoustics, in the theory-forming and evaluation of the dynamic behavior of instruments. The other is the simulation of musical instruments (Essl, 2002, pg. 14).

Some examples of this include Finite Difference Models, Mass-Spring-Damper Networks, and Transmission-Line Methods. While Finite Element Methods are commonly found in research and academia, they are not widely used in industry. This is most likely due to the fact that they are mathematically much more complex than Digital Waveguides, and they tend to not be as consistent either (Karjalainen, 2004, pg. 11). The inconsistency in results comes from the dual delay feedback that can potentially produce divergent results from similar input data.

#### 2.1.2.1 Finite Difference Models

One of the early approaches to digital simulation of physical systems was the finite-difference method. Finite difference models have been used in a couple of different settings with regards to physical modeling. One way is in the study of acoustics, specifically evaluating the dynamical behaviors of musical instruments. Piano-strings, bar percussion instruments, square plate instruments, and the kettledrum are some of the instruments that have been studied using finite difference methods (Essl, 2002, pg. 14).

The other situation is actual simulation of musical instruments. Finite differencing of string equations has been studied for many years, and was used to construct the first known digital model of vibrating strings in the 1970's (Smith, 2006, Finite Difference Methods). Interest in finite element methods saw an increase in the mid-1990's in the context of a one-dimensional wave-equation (Essl, 2002, pg. 14).

Finite difference models are done by replacing derivatives in physical systems with finite differences. There are a couple ways this has been accomplished.

One way consists of an approximation of the first partial derivative with respect to time

$$y(t,x) \approx [y(t,x)-y(t-T,x)]/T \approx [y(t+T,x)-y(t-T,x)]/2T. \tag{2-2}$$

In this equation, $t$ represents time in seconds, $x$ is the position along the string, and $T$ is the sampling period.

A variation is called centered finite difference. It requires an extra factor of two over sampling in its magnitude response for a given accuracy; however it holds the advantage of not introducing a time delay (Smith, 2006, Finite Difference Methods). Centered finite differencing is the most common form of finite difference models. The centered finite difference approximation to the spatial partial derivative *y(x)* is given by

$$y(x) \approx (y(x + \Delta x,t) - y(x - \Delta x/2,t))/\Delta x \qquad \textbf{(2-3)}$$

where $\Delta x$ is the spatial sampling interval (Karjalainen, 2004, pg. 7).

Matti Karjalainen discusses the modeling of one-dimension stringed instruments using finite difference time domain (FDTD) formulations in his paper *1-D digital Waveguide Modeling for Improved Sound Synthesis* (Karjalainen, 2002, pp. II/1869-II/1872). Karjalainen says that modeling using finite difference models is "efficient enough to run in real time on modern processors and more flexible than the computationally less expensive commuted synthesis" (Karjalainen, 2002, pg. vII/1869). Commuted Synthesis is an extension of digital waveguide synthesis used for synthesizing strings (plucked, bowed, and hammered) where the string and resonator portions of the model are commuted (this can be accomplished because they are linear and time-invariant). The excitation portion of the model is then convolved with the resonator impulse response to form a single excitation table (Smith, 2006, Commuted Synthesis of Strings), and (Smith, van Dunye, 1995, pg. 1).

FDTD waveguide structures are being analyzed with various combinations of lossless and lossy propagation, input and output ports, terminations and scattering junctions. Scattering junctions are equations used for ports that have different resistances (Smith, 2006, Scattering Junctions). Karjalainen clearly explains what is available for this technique, and shows the potential it has for accurately modeling stringed instruments. However, finite difference models are not the focus of this research because they are not as common, nor as simple as digital waveguides.

#### 2.1.2.2 Mass-Spring-Damper Networks

The vibrating nature of strings can be modeled using a series of springs and masses. Mass-Spring computational models have been used for years in physics and acoustics for demonstrating vibrations and waves. This technique is efficient in modeling vibrating media because it captures both density and elasticity. The mass-spring computational model has been extended to vibrating surfaces, such as drums, and volumes. This can be accomplished through a fabric of masses connected by multiple springs, arranged in a circular shape (for a drum), or in a lattice shape (for a volume) (Roads, 1996, pp. 271-273). For the purposes of this research project, Mass-Spring-Damper networks are not the focus as they are not as common as Digital Waveguides in industry, or in academia.

#### 2.1.2.3 Transmission-Line Methods

Transmission-Line Methods consist of methods used to solve differential equations, such as highly scattering digital waveguide simulations, digital waveguide meshes, and scattering networks.

Transmission-Line Methods are commonly used in speech-production modeling, but are also used in physical modeling synthesis, and are commonly known as digital waveguide networks, or digital waveguide meshes (Karjalainen, 2004, pg. 6).

Digital Waveguide Meshes came about as a possible solution to enable two and three dimensional FDTD simulations to be run in real-time (Erkut, 2002, pg. 23). The basic building blocks of Digital Waveguide Meshes are paired delay lines for transporting wave signals in opposite directions, and scattering junctions to connect them together (Bilbao, 2004, pp. 12-15). Element ports are connected to the scattering junctions. These then form the mesh-type structures, which include two-port delay lines between nodes (Karjalainen, 2004, pg. 4). Meshes are commonly used in modeling membranes, such as drum heads, and cymbals, as well as three-dimensional rectangular or square shaped objects, such as a cajon, or even a room (Karjalainen, 2004, pg. 17), (Savioja, Valimaki, 1999, pg. 1), and (Van Duyne, Smith, 1999, pg.1).

Transmission-Line Methods are not as common in industry as other physical modeling algorithms, particularly in the United States (Bilbao, 2001, Abstract), and therefore will not be used in this research.

### 2.1.3 Modal Synthesis Methods

Modal synthesis methods concentrate on modeling the spectral responses of physical systems (Essl, 2002, pg. 10). It relies on the fact that any sound-producing object can be represented as a collection of vibrating substructures characterized by modal data (Pakarinen, 2004, pg. 3). Some of these methods include additive sinusoidal modeling, and resonant filter modeling. Modal synthesis methods have been used to model plates and drums (Cook, 2000, pg. 20), but still are not as common as digital waveguides, or even finite element methods, and therefore are not used in this research.

#### 2.1.3.1 Additive Sinusoidal Modeling

Another important sampling technique that Cook discusses is additive synthesis. This is defined as "synthesis of signals by adding fundamental waveform components" (Cook, 2002, pg. 25). Because any time-varying signal can be represented as a linear combination of sinusoidal components, Fourier analysis allows the analysis and resynthesis of waveforms. When only a few sinusoidal components exist, sinusoidal oscillators or resonant filters can be added to model the individual modes. This is also sometimes referred simply as modal synthesis (Cook, 2002, pg. 26).

Another form of additive synthesis is the reassigned bandwidth-enhanced additive model. Kelly Fitz and Lippold Haken discuss this sound modeling technique in their paper *On the Use of Time-Frequency Reassignment in Additive Sound Modeling* (2002, pp. 879-893). This technique produces a sharper, more robust additive representation of the sample. The sample is robust in the sense that it can accurately model many different sound or tone parameters. The reassigned bandwidth-enhanced additive model can handle many different modeling specifications at the same time while still maintaining high sonic integrity.

> "The Reassigned Bandwidth-Enhanced Additive Model follows ridges in a time-frequency analysis to construct partials having both sinusoidal and noise characteristics. This model yields greater resolution in time and frequency than is possible using conventional additive techniques, and better preserves the temporal envelope of transient signals, even in modified reconstruction, without introducing new component types or cumbersome phase interpolation algorithms" (Fitz, Haken, 2002, pg. 879).

The reassignment bandwidth-enhanced additive sound model is similar to traditional models in that a waveform is modeled as a collection of components called partials. The difference is that the reassigned model combines sinusoidal energy and noise energy into a single waveform component having time-varying amplitude, frequency, and bandwidth parameters. The time and frequency estimates used to define the partial parameter envelopes are improved, thereby improving the time-frequency resolution of the representation, and improving its phase accuracy (Fitz, Haken, 2002, pg. 880). The combination of time-frequency reassignment and bandwidth enhancement produces a model having a single component type that is capable of representing at high fidelity a wide variety of sounds, including non-harmonic, polyphonic, impulsive, and noisy sounds. Fitz and Haken effectively and thoroughly cover the reassignment and bandwidth enhancement technique in their paper. However, due to this technique not being as commonly used as digital waveguide methods, it will not be used in this study.

#### 2.1.3.2 Resonant Filter Modeling

Resonant Filter Modeling is a real-time technique used to model the modes of acoustical systems (O'Brien, Cook, Essl, 2001, pg. 2). Resonant filters can be used directly to model formants (Cook, 1998, pg. 4). There has not been as much research conducted on resonant filter modeling, and therefore it will not be used in this research.

### 2.1.4 Banded Waveguides

Cook discusses another sampling technique that is a spin-off of wavetable synthesis in his paper *Theory of Banded Waveguides* (2004, pp. 37-50). Banded Waveguide Synthesis was originally developed for the synthesis of bowed bar percussion instruments (Essl, 2003, pg. 2), and (Essl, Cook, 2000, pp. 1-10). Banded waveguides are "a way of synthesizing sounds made by solid objects and an alternative method for treating two- and three-dimensional objects" (Cook, 2004, pg. 37). It fits in with other physical models' synthesis algorithms. Physical modeling of musical instruments is a synthesis technique that is well established in computer music technology (Cook, 2004, pg. 37). Physical models are historically related to computationally expensive algorithms from the 1960's, but have become more efficient in the last couple years with faster methods such as waveguide synthesis. Digital waveguide models provide discrete-time models of distributed media such as vibrating strings, bores, horns, and plates (Cook, 2004, pg. 37). Banded waveguides are a newer physical modeling technique, and therefore are not as common in industry and academia as digital waveguides at the time of this study. For this reason they will not be used in this research.

### 2.1.5 Combination Systems

In a paper entitled "Digital Waveguides vs. Finite Difference Schemes: Equivalence and Mixed Modeling (2004)", Karjalainen and Erkut discuss a method used to construct mixed models by combining Finite Difference physical elements and Digital Waveguide wave elements through a converter. There are two ports, with a Finite Difference junction on one end and a Digital Waveguide junction on the other (Karjalainen, 2003, pp. 4-5). The converter maps the physical variables of the Finite Difference junction (for example, string displacement, velocity, etc.) to a wave port of the Digital Waveguide junction, and vice-versa (Karjalainen, 2004, pp. 11-12), and (Smith, 2004, pg. 1). An example of a mixed system could be a drum membrane, where a rectangular Finite Difference waveguide mesh could be used with Digital Waveguide elements at the boundaries (Karjalainen, 2004, pg. 17). There has also been a small amount of research conducted toward combing physical modeling synthesis with other audio synthesis techniques (Valimaki, 1995, pg. 15).

## 2.2 Perceptual Quality Measurement Systems

The quantitative measurement of the perceptual quality of sound simulations is a research area that hasn't been explored very much. However, there have been recent studies of perceptual models in the areas of digital speech coding (for wireless telephony) (Beerends, Hakstra, Rix, Hollier, 2001), as well as in music coding for compression (Essl, 2002, pg. 128). Some models have also been associated with audio quality testing of MP3 (Painter, Spanias, 2000, pp. 27-28). Some of these models are the Perceptual Audio Quality Measure (PAQM), the Noise-to-Mask Ratio (NMR), the Disturbance Index (DIX), the Objective Audio Signal Evaluation (OASE), the Perceptual Evaluation (PERCEVAL), the Perceptual Objective Measurement (POM), and the toolbox approach.

### 2.2.1 Perceptual Audio Quality Measure (PAQM)

The PAQM is different than other audio quality measurement systems in that it doesn't characterize the actual audio system being tested, it characterizes the perception of the system's output signal. A comparison is performed between the degraded output and the ideal reference. The comparison utilizes a model of the human auditory system, allowing predictions to be made about the subjectively perceived audio quality of the system output using an input signal. The internal representation of the reference and the degraded output is calculated using a perceptual model. The difference between the reference and the degraded signal is obtained by subtracting them from one another. An interpretation of the difference of these internal representations is then performed with a simple cognitive model (Beerends, 1998, pp. 1-37). The PAQM has been tested in applications of both wideband music codes and telephone-band speech codecs, and has recently been adopted as an internal reference by the ITU-R (International Telecommunication Union – Radiocommunication group) for analysis of wideband codes in the 20Hz to 20kHz range (Perceptual Audio Quality Measurement, 2005).

### 2.2.2 Noise-To-Masked-Ratio (NMR)

Noise-To-Masked-Ratio (NMR) is very similar to PAQM, with even the exact same test procedures performed on the reference signal (Thiede, Kabot, 1996, pg. 11). NMR evaluates the level-difference between the masked threshold and the test signal. A DFT is used to analyze the frequency content of the signal. To represent pitch perception, a scale transformation from frequency to pitch is done using the Bark scale. The pitch scale resolution is about 1 Bark. NMR doesn't require lots of computational power, so it was possible to implement NMR as a real-time system early in its development. NMR has been in use since 1987, and has proven its basic reliability, specifically in the area of telephony (International Telecommunication Group, 2001, pg. 28).

While the PAQM uses smearing and compression to model the masking of the human auditory system on both signals, the NMR only performs these operations on the original signal. The PAQM measures predicted perceptual differences between both signals, while the NMR measures the level difference between the masked signal (original) and the degraded, or noisy signal (test signal).

### 2.2.3 Disturbance Index (DIX)

Disturbance Index (DIX) is a perceptual measurement method based on an auditory filter bank which yields a high temporal resolution, provides precise modeling of masking, and a refined analysis of the envelopes within each filter channel (Thiede, Kabot, 1996, pg 1).

The center frequencies of the filters are distributed over a pitch scale. The top of the filter shape is rounded to make sure that the number of filters converts the frequency range without ripples in the frequency response. To model masked thresholds, the filter slopes decrease exponentially over the Bark scale. The steepness of the slopes depends on the level of the input signals (International Telecommunication Group, 2001, pp. 27-28).

While the filter bank approach of the DIX system generally yields slightly better temporal resolution than FFT-based systems, it is also much more time-consuming. For this reason it is not as common as other perceptual measurement systems, and therefore will not be used in this research study.

### 2.2.4 Objective Audio Signal Evaluation (OASE)

Objective Audio Signal Evaluation (OASE) is a perceptual measurement system that is similar to the DIX system outlined in the section above. It also uses a filter bank to analyze input signals. It is a much more comprehensive system, using four times as many filters and spacing the frequencies six times closer together over the pitch scale (International Telecommunication Union, 2001, pp. 28-29).

While the OASE system is very accurate even when compared to the DIX system, the complexity also makes it much slower, and not as prominent as some other perceptual systems. Therefore the OASE system will not be used in this thesis.

### 2.2.5 Perceptual Evaluation (PERCEVAL)

Perceptual Evaluation (PERCEVAL) is a perceptual evaluation technique first introduced in 1992. It laid the foundation for the PAQM mentioned above in section 2.2.1. Like the PAQM, the PERCEVAL performs a Fourier transform, uses a transform for weighting of the human ear, calculates a frequency-to-pitch transformation, and uses convolution for spreading (International Telecommunication Union, 2001, pg. 30).

The PAQM essentially replaced the PERCEVAL when it was introduced six years after the PERCEVAL system. The PAQM slightly changed the transform used for the frequency to pitch transformation, and changed the way the results were processed resulting in an actual quality measure score. The PERCEVAL results only show the probability of detecting distortions (or differences) between the two signals. This thesis will not test the PERCEVAL, as the PAQM was chosen as one of the perceptual methods to test, and it is an updated, more current test than the PERCEVAL (International Telecommunication Union, 2001, pg. 30).

#### 2.2.6 Perceptual Objective Measurement (POM)

Perceptual Objective Measurement (POM) is another perceptual test that is very similar to the PAQM and PERCEVAL algorithms mentioned above. The purpose of the POM is to quantify the amount of degradation that might occur between a reference signal and its "degraded version" (International Telecommunication Union, 2001, pp. 30-31). Again this is accomplished by the same steps as the PAQM and PERCEVAL, however the transform modeling the human ear is performed first, before the Fourier transform. POM uses a detection process to determine the amount of degradation (or difference) between the reference signal and the degraded (or test) signal. The POM went a step forward by not only resulting in a probability of detecting a difference between the two signals, but also resulting in an actual distance representing the perceptual gap between the two signals (International Telecommunication Union, 2001, pp. 30-31).

The POM will not be used in this study as it is not as recent or common as the PAQM and NMR tests in industry and academia.

#### 2.2.7 The Toolbox Approach

The toolbox approach is a perceptual quality technique that measures the perceived distance in audio quality of an audio signal in relation to a test signal. This results in an indication of the overall subjective audio quality level of the test signal (International Telecommunication Union, 2001, pg. 31).

There are three main steps to the toolbox approach. The first is the calculation of the loudness. This is performed using a large FFT of 2048 points. Other parameters including sharpness, the amount of pre-echoes, and masked loudness are calculated as well. The second step includes weighting procedures that depend on the amount of the perceived loudness difference, and the variation of loudness in time. The final step of the toolbox approach includes the generation of a set of output values based on a statistical analysis of the values from the previous steps. These values include the mean, maximum, root mean square, and the standard deviation of the mean. A weighted sum of these intermediate values is used for the final fitting of the distance between the two signals being tested (International Telecommunication Union, 2001, pg. 31).

While the toolbox approach certainly fits the purpose of this research in finding an adequate perceptual testing method for synthesized audio, it is a slower method, and it is rare in academia and industry, and this research did not uncover much information about it in comparison to some of the other test methods listed above. Therefore the toolbox approach will not be used in this research.

## 2.3 Subjective Listening Test Method

It is not possible to validate an objective test directly. Therefore the objective test methods must be validated against subjective listening tests.

A common subjective listening test in industry and academia is the Mean Opinion Score (MOS). There are several characteristics of the MOS tests. At least six listeners (preferably at least twelve) are selected at random (International Telecommunication Union, 1996, pg. 27). The listeners participate in the listening tests in the same room all together at the same time (International Telecommunication Union, 1996, pg. 14). The samples being tested are then played back to back, and the listeners rate the quality, or similarity of the samples based on a scale of one to five (International Telecommunication Union, 1996, pg. 18). The test is repeated various times at different listening levels or conditions (noisy background, extreme quiet, etc.) (International Telecommunication Union, 1996, pp. 18-20). The numerical mean is then taken for each test at each listening level. This mean is the actual MOS score. Further analysis can be carried by analysis of variance using the different listening levels. Calculation of the standard deviation for each condition is not recommended (International Telecommunication Union, 1996, pg. 20).

The MOS test is a very common subjective listening test, especially in the area of telephony. Other audio quality subjective tests have been studied (Bonebright, Miner, Goldsmith, Caudell, 1998, pp. 1-8), and (Van del Doel, Pai, Adam, Kortchmar, Pichora-Fuller, 2002, pp. 1-6). However, the MOS is a standard which was set by the International Telecommunication Union in 1996. Elements of the MOS test will be used with the subjective listening tests executed in this research.

## 2.4 Conclusions

While there is still work to be done, there has been much research conducted in the area of acoustical physical modeling. The areas of digital waveguide synthesis, finite element methods, modal synthesis methods, banded waveguides, and combination systems have all been studied, and many have been put into use in industry as well. Chapter 3 includes an industry survey showing that digital waveguide synthesis is the most common physical modeling technique used in industry, justifying its use as the exclusive form of physical modeling used in this thesis.

There have also been many studies conducted in the area of perceptual audio quality, specifically in speech coding and music coding for compression. Many perceptual test methods, including the PAQM and NMR algorithms, have been tested and set as audio quality test standards by the ITU-R.

However, there haven't been studies conducted to show whether perceptual audio quality tests like the PAQM and NMR are good measurement techniques for the quality of physical modeling synthesis. The purpose of this research is to determine if the PAQM and/or NMR quality tests can effectively predict the audio quality in the physical modeling technique of digital waveguide synthesis when compared against a subjective listening test performed by a qualified panel of listeners.

# 3 Methodology

## 3.1 Introduction

Research for this thesis began with an industry survey which reported that Digital Waveguide synthesis is the most common form of physical modeling synthesis currently in use today. Following the survey reporting, it was the purpose of the study to obtain samples for testing; acoustic samples are recorded directly from an acoustic guitar. The guitar sample was processed through a soundhole pickup. Each of the synthesized samples was obtained using a Karplus-Strong algorithm digital waveguide program controlled via MIDI (Musical Instrument Digital Interface).

After completing the sample measures, both the PAQM (perceptual audio quality measure) and NMR (noise-to-mask ratio) algorithms were created and executed in a Matlab environment, testing and generating a comparative score between corresponding pairs of acoustic and synthesized samples. A corresponding subjective listening test was then administered to a listening panel resulting in quality measure scores for each sample pair, based on the same format used in the Mean Opinion Scores (MOS). Finally a comparison was performed using the results from the computer tests, and the means and standard deviations from the listening test results.

### 3.2 Synthesizer Industry Survey

The survey was completed using some of the major synthesizer manufacturers to determine which form of physical modeling is currently used the most, in order to determine which method would be most common for testing purposes. Yamaha, Korg, Roland, Alesis, Clavia, Casio, Kurzweil, and Moog were all contacted via email, phone, or personal interviews. The purpose was to determine if their keyboards contained physical modeling synthesis technology, and if so, which specific techniques are used most frequently. Many companies were unable to provide specific details due to trade secrets. However, generally enough information was either given or already publicly known in order to draw conclusions as to the technology behind the sounds.

#### 3.2.1 Yamaha

Yamaha utilizes several different proprietary physical modeling methods. Some of these include Self-Oscillating Virtual Acoustic Modeling, Analog Physical Modeling Synthesis, Formulated Digital Sound Processing, and Virtual Circuit Modeling (B. Angelos, personal communication, February 6, 2006). Yamaha has had a long standing relationship with Stanford University, home of the Computer Center for Research in Musical Acoustics (CCRMA). In fact, these two organizations hold many joint patents. Julius O. Smith of Stanford developed the Virtual Acoustic Modeling technology used by Yamaha, which included the work he had done with Digital Waveguide Synthesis. Many of the Digital Waveguide files found in the open-source Synthesis Toolkit collection state that the patents are held by Yamaha. Therefore, from the information gathered above, it can be assumed that Yamaha incorporates Digital Waveguide Synthesis technology in their synthesizer division. See appendix A for source code.

### 3.2.2 Korg

While Korg did not respond to survey requests, the following research shows that it is likely that Korg uses Digital Waveguide Synthesis, due to their strong ties to Yamaha. Yamaha owned a controlling interest in Korg's stock from 1989 to 1993, and was a major partner before and after that time period, supplying Korg with circuitry and parts. Also in 1994 Korg licensed Yamaha's Virtual Acoustic physical modeling technology (Russ, 1994). This license was renewed again with Yamaha and Stanford in 1998 (Sondius-XG, 1999).

### 3.2.3 Roland

The significance of sound test properties processed through Roland would appear to be based on unique and specialized test measures available only to Roland. It is important to recognize that Roland, although proprietary, does use Physical Modeling (J. Gardner, personal communication, February 9, 2006).

### 3.2.4 Alesis

Alesis uses physical modeling technology in their Fusion synthesizers. Alesis was unable to provide specific details as to what type of physical modeling synthesis they use, only that their reed and pipe sounds are the only sounds to use physical modeling (R. Greenly, personal communication, February 3, 2006). Reeds and pipes could potentially use either digital waveguides or finite element methods.

### 3.2.5 Clavia

Clavia is most famous for their Nord synthesizers. Clavia confirmed they use physical modeling synthesis, however they were unable to provide details as to what type of techniques they use (J. Smith, personal communication, March 9, 2006). The Nord Modular G2 is known to primarily use Digital Waveguide Synthesis, along with Modal Synthesis. The Nord Modular G2 uses Digital Waveguide Synthesis with blown pipes, flutes, reed woodwinds, brass, pipe organs, and bowed strings (Singer, 2004).

### 3.2.6 Casio

Casio did not provide any response to inquiries regarding their keyboards. Excluding their high-end digital pianos, most of their synthesizers are relatively inexpensive. It is therefore highly unlikely that Casio uses any physical modeling synthesis techniques, and most likely uses sampling technology.

### 3.2.7 Kurzweil

Kurzweil only uses physical modeling technology in their KB3 organ modes, which can be found in the K26, K2661, PC2, and K25 (which has been discontinued) synthesizers (J.R. Bellefeuille, personal communication, February 3, 2006). Organ sounds are typically modeled using Digital Waveguide Synthesis.

### 3.2.8 Moog

Moog synthesizers do not use any physical modeling technology. In fact, they don't use any digital technology as all the synthesis is done via analog subtractive synthesis (A. Gaynes, personal communication, February 15, 2006).

### 3.2.9 Industry Survey Conclusion

The survey shows that sixty-six percent of the companies who reportedly use physical modeling synthesis use waveguide synthesis as the primary form. Due to the predominance of digital waveguide synthesis, the decision to pursue this technology for producing test samples for this research has been justified. Therefore waveguide synthesis will be the audio physical modeling synthesis technique used in this thesis.

## 3.3 Test Samples

Synthesized and acoustic samples were needed for this research. To collect them, musical notes from an acoustic guitar were recorded, and physical modeling synthesis software was used for the synthesized samples. The following sections detail these procedures.

### 3.3.1 Hardware Configuration

The USBOmni Studio from M-Audio was the USB interface used to capture the original acoustic samples as well as the MIDI information used to control the synthesized sounds. A Casio CTK-541 was used as a MIDI controller to transmit and control the MIDI information.

### 3.3.2 Recording Software

Cubase SX 2.0 by Steinberg was used to capture both the acoustic and the synthesized samples. Matlab 6.5 was used to do all signal processing.

### 3.3.3 Original Samples

Original plucked string samples were obtained by recording an Epiphone dreadnought acoustic guitar at a 44.1 kHz sampling rate, and at a 16-bit resolution in the Cubase program. Single notes were recorded direct using a Fender soundhole transducer pickup. All seven notes were recorded from a C major diatonic scale, which includes 261.6 Hz, 293.7 Hz, 329.6 Hz, 349.2 Hz, 392 Hz, 440 Hz, and ends at 493.8 Hz.

### 3.3.4 Digital Waveguide Samples

MAX/MSP software from the Cycling '74 company was used to create the samples of Digital Waveguide Synthesis. The Karplus-Strong algorithm was the Digital Waveguide technique employed, and the sounds were controlled via MIDI using the same Casio keyboard mentioned in section 3.3.1. Again, all seven notes from a C major diatonic scale, beginning at 261.6 Hz were recorded for comparison.

## 3.4 Perceptual Audio Quality Measure (PAQM)

The following sections will describe the steps needed to perform audio quality testing by way of the PAQM. These steps include performing a Fourier transform, passing the samples through a transfer function representing the human ear, transforming the samples from frequency to pitch, time-frequency domain smearing, and comparisons between the final versions of the acoustic and synthesized samples. The final result is a quality measurement score between 1 and 5. A lower score indicates that the original (acoustic) and test (synthesized) samples are very similar, while a higher score would indicate that the two samples are very different from one another. See appendix B for source code.

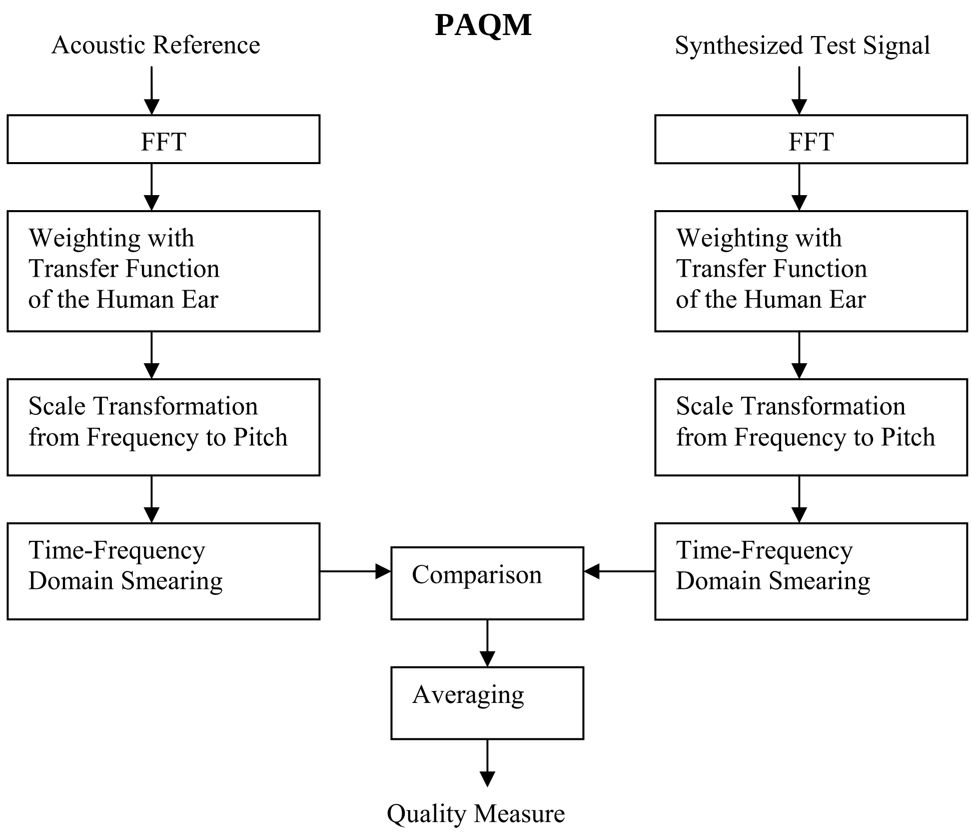


**Figure 3-1: PAQM (Thiede, Kabot, 1996, pg. 11)**

### 3.4.1 FFT

The first step in the PAQM is to perform a Fourier Transform on both the original and synthesized samples for a time-frequency mapping. In order to accomplish this, an FFT is performed using the built-in function in Matlab.

### 3.4.2 Weighting with Transfer Functions of the Human Ear

The next step is to model hearing of the human ear via intensity warping. This is accomplished by passing both the acoustic and synthesized samples through a transfer function that represents the outer and inner ear. This also serves as a form of compression on the samples. In 1979, Ernst Terhardt published the following equation which represents the transfer function from outer ear to inner ear:

$$W[k]/\mathrm{dB} = -0.6 \cdot 3.64 \cdot \left(\frac{f[k]}{\mathrm{kHz}}\right)^{-0.8} + 6.5 \cdot e^{-0.6 \cdot \left(\frac{f[k]}{\mathrm{kHz}} - 3.3\right)^2} - 10^{-3} \cdot \left(\frac{f[k]}{\mathrm{kHz}}\right)^{3.6}$$

**Figure 3-2: Terhardt's inner ear transfer function (International Telecommunication Union, 2001, pg. 49)**

where $f[k]$/kHz represents the frequency being passed through the transfer function. Passing both the acoustic and synthesized samples through this transfer function represents the weighting of the human ear.

### 3.4.3 Scale Transformation from Frequency to Pitch

The next step in the PAQM system is to convert the sample from its actual frequency value to the pitch at which it is heard. This is sometimes referred to as frequency warping (Esquef, 2004, pg. 7).

The frequency is converted to the Bark scale by way of the following equation, first published by Hartmut Traunmuller in 1990:

$$Criticalbandrate(bark) = [26.81/(1+1960/f)]\text{-}0.53 \qquad \textbf{(3-1)}$$

Passing both the acoustic and synthesized samples through this conversion will successfully transform the values from frequency to pitch.

### 3.4.4 Time-Frequency Domain Smearing

Time-frequency domain smearing (or spreading) can be accomplished through non-linear convolution. For both acoustic and synthesized samples, convolving the scaled samples from the previous step with the compressed versions from the second step allows modeling of the masking behavior of the human auditory system.

### 3.4.5 Comparison/Averaging

Finally a comparison is made between the acoustic and synthesized samples by taking the difference between them. The absolute value of the difference is calculated, and then divided by the sample size for averaging. This value is then rounded, finally ending with the quality measure score.

## 3.5 Noise-to-Mask Ratio (NMR)

The following sections describe the steps for administering the NMR test. These steps include performing a Fourier transform on the samples, passing the samples through a transfer function representing the human ear, transforming the samples from frequency values to pitch values, time-frequency domain spreading on the acoustic sample, and finally performing a comparison between the final versions of the samples. The final result is a quality measurement score between 1 and 5. As with the PAQM, a lower score indicates that the two samples are very similar, while a higher score indicates that the two samples are very different from one another. See appendix C for source code.

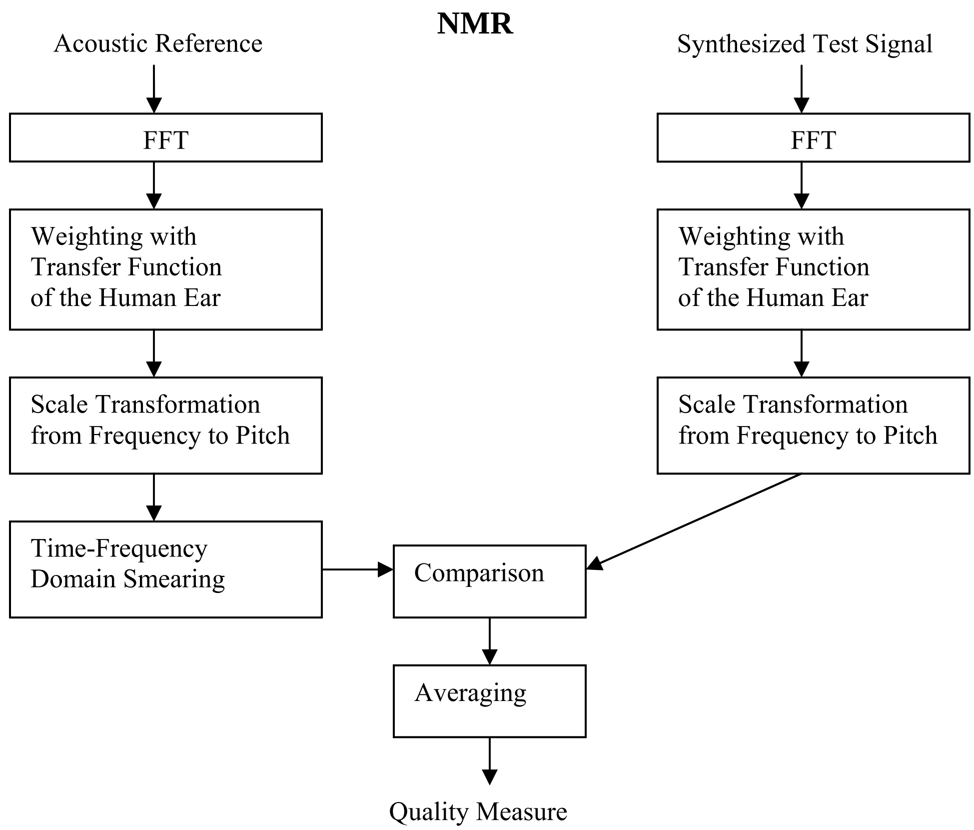


**Figure 3-3: NMR (Thiede, Kabot, 1996, pg. 11)**

### 3.5.1 FFT

Like the PAQM, the first step in the NMR is to perform Fourier Transforms on both the original and synthesized samples. An FFT performed in Matlab accomplished this task. This transformed the signals from the time domain to the frequency domain in order to work with the frequency content of the signals.

### 3.5.2 Weighting with Transfer Function of the Human Ear

Again like the PAQM, the next step is the intensity warping to model the hearing of the human ear. Both acoustic and synthesized samples are passed through the transform based on Terhardt's equation (see figure 3-2), compressing the samples in the process.

### 3.5.3 Scale Transformation from Frequency to Pitch

Just like the PAQM, the next step in the NMR is to convert both the acoustic and synthesized samples from their frequency values to their pitches. Passing the samples through Traunmuller's equation shown in fig. 3-3 accomplished this task. This is the last step before the comparison/averaging for the synthesized sample.

### 3.5.4 Frequency-Time Domain Smearing

Non-linear convolution is performed on the acoustic sample to accomplish the frequency-time domain smearing. The scaled acoustic sample from section 3.7.3 is convolved with the compressed acoustic sample from section 3.7.2. This models the masking characteristics of the human auditory system on the original, acoustic sample. This step is not performed on the synthesized test signal/sample, as the NMR tests the level difference between the masked signal (original) and the degraded/noisy signal (synthesized).

### 3.5.5 Comparison/Averaging

To compare the acoustic and synthesized samples, the difference is taken between the acoustic sample from the previous section and the synthesized sample from section 3.7.3. The absolute value is taken of the result, and then divided by the total number of samples. The result is rounded, and the final result is the quality measure score.

### 3.6 Listening Tests

Listening test surveys were distributed to a panel of listeners. Unlike the MOS test where the panel is comprised of random individuals, a qualification was established in order to have a more musically experienced listening panel. The qualification to participate in the listening test was that the participant had to have been previously or currently be enrolled in a music course at a high-school level or above. This produces a more critical listening panel than a group of random individuals would. It was decided that the number of participants used in the listening test would be twenty, which is in accordance with the Mean Opinion Score (MOS) format (International Telecommunication Union, 1996, pg. 27). The participants' ages ranged between nineteen and twenty-nine years old.

CD's were distributed containing all of the acoustic and synthesized samples. Unlike the MOS, the participants were instructed to listen to the samples on the best listening device that they were most accustomed to using. It was decided this would be better than having the listening group all listen to the samples together at the same time on an audio system or device with which they were not familiar. The CD would play a note from the C major diatonic scale from either the acoustic or the synthesized sample, followed by the same note from whichever type of sample wasn't played first.

Survey sheets were also distributed along with the CD's (see appendix D). The sheet contained a place for each note from the C major diatonic scale where the participants would score how similar or different the two samples sounded. A likert scale of 1-5 was given for the listeners to use, starting with a 1, meaning the tones sounded exactly alike, and ending with a 5, meaning the tones sounded nothing alike. Following the MOS format (International Telecommunication Union, 1996, pp. 19-21), the average was taken for each note from the surveys resulting in a final likeness score for each note in the scale.

### 3.7 Comparing the Subjective Listening Test Scores against the Objective PAQM and NMR Test Scores

Using the data obtained from the subjective listening tests, the standard deviation can be computed for each sample using the MOS score for the average. The scores from the PAQM and NMR tests can then be analyzed to see if their respective scores fall within one standard deviation of the corresponding MOS. As a consideration to accept subjective values of tonal measures, a range of one standard deviation is acceptable. If the objective test scores fall within one standard deviation, the test method is defined to be valid. Otherwise the objective test methods will be defined as unacceptable test methods for digital waveguide synthesis. This procedure is repeated for each sample.

# 4 Results and Analysis

## 4.1 Introduction

This chapter will discuss those results from the procedures outlined in chapter 3. A comparative analysis will be conducted which will show whether or not the PAQM and/or NMR testing procedures are valid test measurements when compared with the results from the listening test.

## 4.2 PAQM

The following are the quality measurement scores from the PAQM test described in chapter 3. As a brief review of this testing procedure, the synthesized signal and an actual recording were each fed into the PAQM test where the following steps were performed on each signal. First an FFT was performed, then the samples were passed through a transform representing the weighting of the human ear, then the values were transformed from frequency to pitch, and then time-frequency domain smearing is executed through convolution. Finally a comparison of the two signals results in the quality measurement scores listed below.

**Table 4-1 PAQM Quality Measurement Scores**

| Music Note | PAQM Quality Measurement Score |
|---|---|
| A | 4 |
| B | 3 |
| C | 3 |
| D | 3 |
| E | 5 |
| F | 3 |
| G | 4 |

A smaller quality measurement score means the two samples are very similar, while a larger score indicates the samples are very different. A smaller score is desirable for a synthesis technique because the idea is to closely match the sound of a real instrument. Following the table above, one can see that samples B, C, D, and F all had a score of 3. This score falls in the middle of the scoring range, and indicates that these sets of samples only sound somewhat similar when compared against one another. Samples A and G received a score of 4, meaning that these sample sets don't sound very similar, and finally sample E received a score of 5, indicating that the E samples sounded nothing alike.

### 4.3 NMR

The following table shows the quality measurement scores from the NMR test described in chapter 3. The test setup was very similar to that for the PAQM algorithm, with the only major difference being that the time-frequency domain smearing is only carried out on the original acoustic sample.

**Table 4-2 NMR Quality Measurement Scores**

| Music Note | NMR Quality Measurement Score |
|---|---|
| A | 3 |
| B | 2 |
| C | 2 |
| D | 2 |
| E | 4 |
| F | 2 |
| G | 3 |

The quality measurement scores for the NMR have the same meaning of the PAQM, as was reviewed in the previous section. For this test, samples B, C, D, and F received a score of 2, meaning that these sets of samples sounded very similar. Samples A and G received a score of 3, indicating that these sample sets only sound somewhat similar. Finally sample E received a score of 4, indicating that the E samples did not sound very similar.

## 4.4 Listening Test

The following table shows the scores from the listening tests described in chapter 3. Again for the listening tests, twenty individuals were given CD's containing tones produced by the original acoustic guitar and a synthesized version using physical modeling synthesis. Two of the same notes were played back to back, one of the notes being the original tone and the other being the synthesized version. The listeners were instructed to rate the two notes based on how similar they sounded. The average was taken for all twenty participants, and these results are shown in the table below.

**Table 4-3 Listening Test Score Averages**

| Music Note | Listening Test Score Averages |
|---|---|
| A | 3 |
| B | 2.75 |
| C | 3.75 |
| D | 2.55 |
| E | 2.8 |
| F | 2.85 |
| G | 3.25 |

Like the PAQM and NMR quality measurement scores, the listening test scores follow the same scoring range format with a lower score indicating great similarities between the two samples, and a higher score indicating great differences.

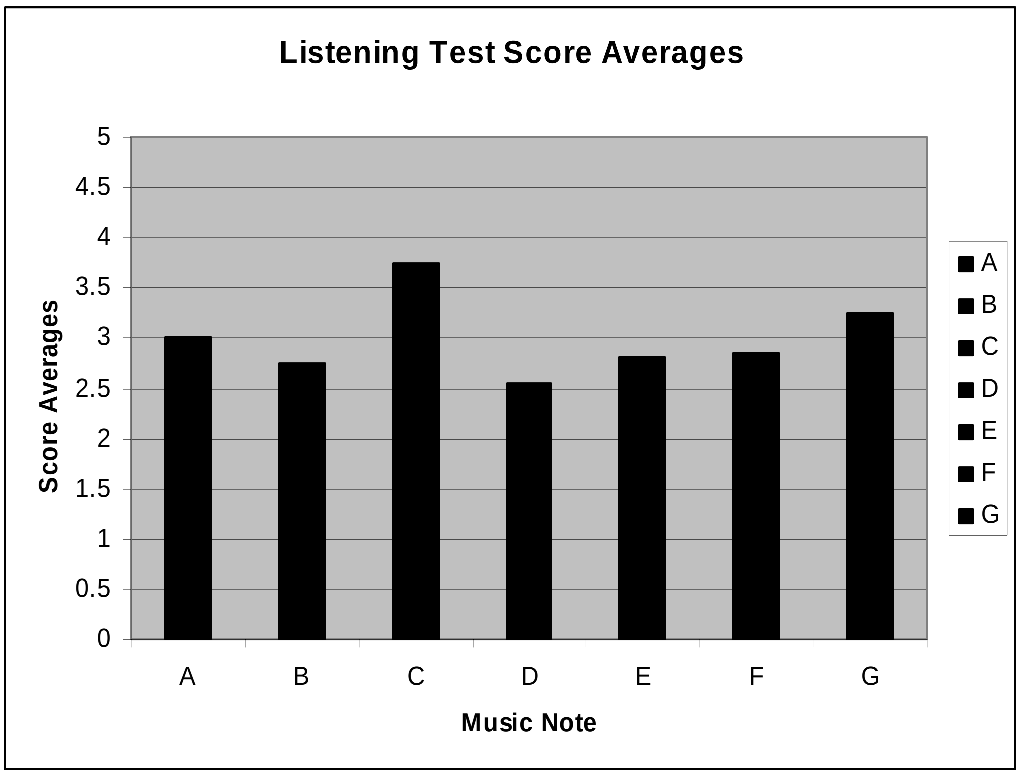


**Figure 4-1 Listening Test Score Averages**

As can be seen in the graph above, outside of sample C, the samples all received similar scores around the middle, indicating that the sample sets sounded somewhat similar. Sample C received a higher mark than the others, meaning that the C samples didn't sound very similar. An analysis of the comparison between the listening test score results and the results from the PAQM and NMR tests is presented in the next section.

As was discussed in section 3.9, the listening panel was made up of twenty individuals (the MOS test prefers at least twelve). The following binomial statistic can be used to test the repeatability of the test:

$$1-(1-p)^n \tag{4-1}$$

where $p$ is the probability of the event reoccurring, and the $n$ represents the sample size. If we want a 95% chance of obtaining the same results with a sample size of 20, then plugging those numbers into the equation comes out to be 100%. In other words, if we repeat the same tests with twenty participants, there is a 100% chance that ninety-five out of one hundred times the results will be the same.

## 4.5 Analysis

The following table shows the standard deviation values taken from the listening test results described in the previous section, as well as in chapter 3. The standard deviation is very pertinent to the testing for a couple reasons. First it shows that the listeners varied slightly and were not completely unified in their assessments of the similarities of the samples under review. Secondly, these slight variations produce the range in which the quality measurement scores from the PAQM and NMR tests need to hit in order to validate their respective results.

**Table 4-4 Standard Deviation Results from Listening Test Data**

| Music Note | Listening Test Score Standard Deviations |
|---|---|
| A | 0.707107 |
| B | 0.766485 |
| C | 0.942072 |
| D | 0.589491 |
| E | 0.748331 |
| F | 0.792149 |
| G | 0.99373 |

From chapter 3, the validation process used to evaluate the quality of the PAQM and NMR tests were to determine if their respective measurement scores fell within one standard deviation from the listening test score averages. The following sections will show if the PAQM and/or NMR are valid objective test measurement systems for digital waveguide synthesis.

### 4.5.1 PAQM results

The following table shows the acceptable ranges for each note. The ranges are defined as being one standard deviation from the listening test score mean. The PAQM quality measure scores are also shown again in order to perform a side-by-side comparison to determine if the results prove the validity of the PAQM.

**Table 4-5 PAQM Acceptability Results**

| Music Note | Acceptable Range | PAQM Score | Acceptable |
|---|---|---|---|
| A | 2.293 – 3.707 | 4 | No |
| B | 1.984 – 3.516 | 3 | Yes |
| C | 2.808 – 4.692 | 3 | Yes |
| D | 1.961 – 3.139 | 3 | Yes |
| E | 2.052 – 3.548 | 5 | No |
| F | 2.058 – 3.642 | 3 | Yes |
| G | 2.257 – 4.243 | 4 | Yes |

While most of the samples tested acceptable, samples A and E did not fall within one standard deviation of the listening test results, and therefore are not acceptable. With only 71 percent of the samples falling within one standard deviation of the listening test scores, the PAQM is defined as not being a valid objective test measurement for a Karplus-Strong algorithm version of the digital waveguide physical modeling form of audio synthesis. The graph shown in figure 4-2 gives another view of the PAQM quality measure scores, and how they compare with the acceptable ranges taken from the standard deviations of the average listening test scores.

An interesting observation from the results is that samples C and G had the largest standard deviation, and therefore the largest range. As would be expected, both of these samples test within the range in the PAQM algorithm. However, sample D had the smallest range, and it also fell within its given range when tested with the PAQM algorithm.

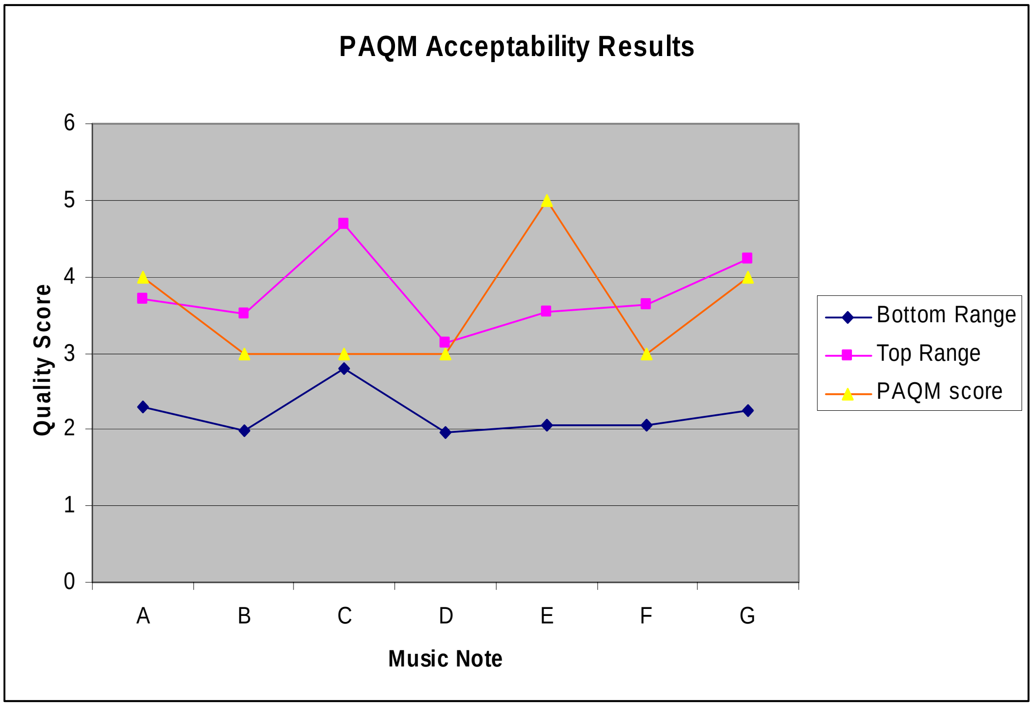


**Figure 4-2 PAQM Acceptability Results**

### 4.5.2 NMR results

The following table again shows the acceptable ranges for each note, with the NMR quality measurement scores listed in the adjacent column. A comparison of the two will prove whether the NMR test is a valid objective test measurement or not.

**Table 4-6 NMR Acceptability Results**

| Music Note | Acceptable Range | NMR Score | Acceptable |
|---|---|---|---|
| A | 2.293 – 3.707 | 3 | Yes |
| B | 1.984 – 3.516 | 2 | Yes |
| C | 2.808 – 4.692 | 2 | No |
| D | 1.961 – 3.139 | 2 | Yes |
| E | 2.052 – 3.548 | 4 | No |
| F | 2.058 – 3.642 | 2 | No |
| G | 2.257 – 4.243 | 3 | Yes |

Samples C, E, and F did not fall within one standard deviation of the average scores from the listening tests, and therefore are not acceptable. Only 57 percent of the quality measurement scores from the NMR test fell within one standard deviation of the listening tests. Therefore the NMR algorithm is also defined as not being a valid objective test measurement for a Karplus-Strong algorithm version of digital waveguide physical modeling audio synthesis. The graph below gives another view of the NMR scores and how they compare against the acceptable ranges listed above.

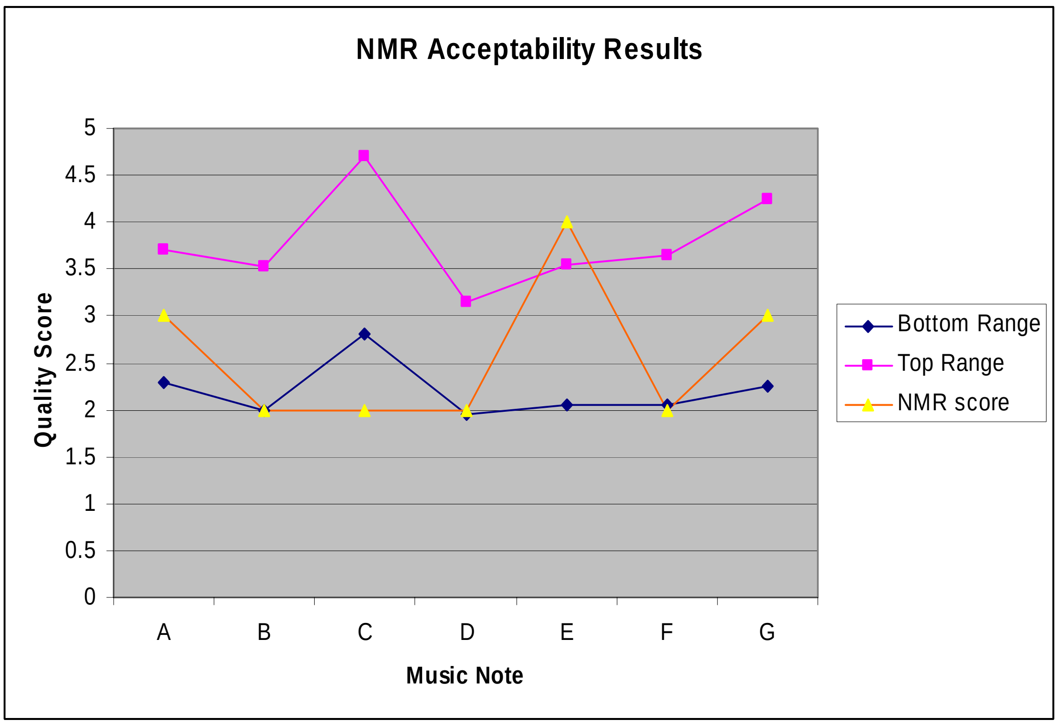


**Figure 4-3 NMR Acceptability Results**

There are some interesting observations to note. Again samples C and G had the largest standard deviations, and therefore the largest ranges within which a NMR quality measurement score would be accepted. However, the NMR score for sample C still fell outside of the range and was one of the three samples that failed to fall within their respective ranges. Sample D had the smallest standard deviation and range, but its corresponding NMR quality score still fell within the acceptable range.

### 4.5.3 PAQM vs. NMR Results

An interesting comparison of the results from the PAQM and NMR tests reveals that the scores from the two tests follow the same pattern as one another. This can be seen clearly in the chart shown below in figure 4-4. The difference is that all of the samples tested with the NMR algorithm scored one mark lower than those of the PAQM. While the PAQM score is based on the predicted perceptual differences between the two signals, the NMR score is based on the level difference between the original masked signal and the test (synthesized) signal. The results below essentially show that the difference between the levels of the masked acoustic samples and the synthesized samples (NMR) is smaller than the predicted perceptual difference between the two samples (PAQM). This indicates that the NMR reported that all sample sets of synthesized and recorded notes sounded more similar to one another than the results from the PAQM.

While a comparison of the results from the PAQM and NMR tests shows a definite relation between the two, a comparison with the standard deviation ranges from the listening tests produced mixed results. Sample E failed on both tests, while sample A failed on the PAQM analysis, but not the NMR. Conversely samples C and F failed on the NMR comparison, but not on the PAQM. Samples B, D, and G were the only samples to fall within the acceptable ranges on both tests. The results also show that the size of the acceptable range does not seem to be a deciding factor, as the sample with the largest range (sample G) fell in the acceptable range for both tests, as did the sample with the smallest range (sample D).

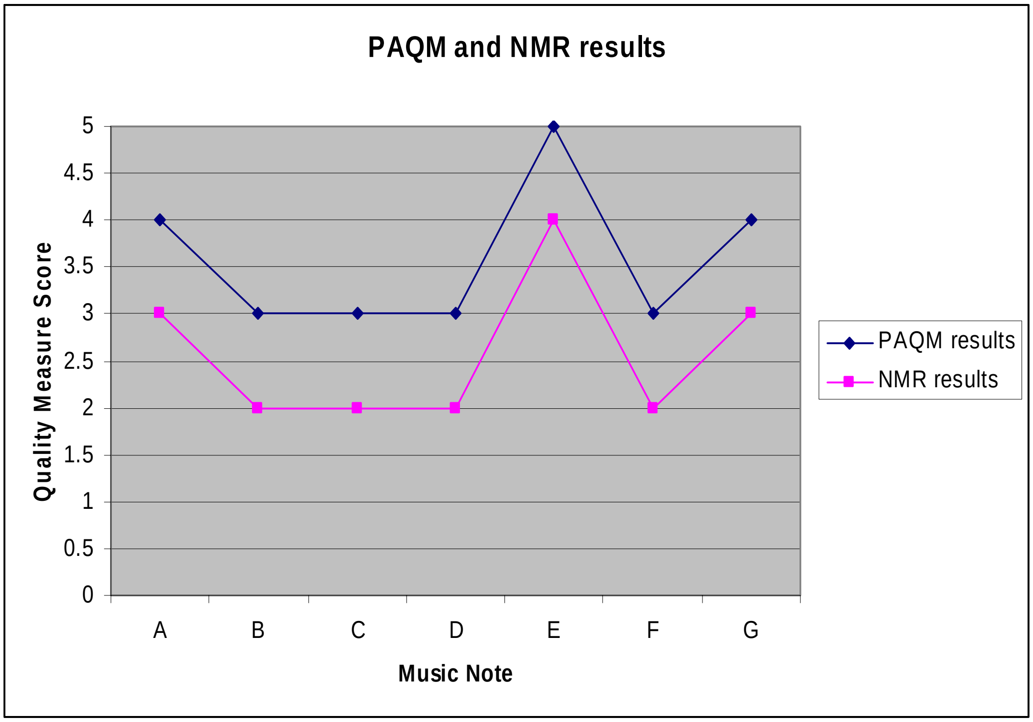


**Figure 4-4 PAQM and NMR results**

# 5 Conclusion and Recommendations

## 5.1 Research Summary

Audio physical modeling synthesis is one of the newest audio synthesis techniques to become commercially available in the last thirteen years. While various physical modeling techniques exist, digital waveguide synthesis has emerged as the most common form, both in industry and in academia.

Audio quality testing has existed for years, with the primary form being subjective group listening tests. While these tests have proven to be sufficiently accurate, they are generally expensive and time-consuming (Penttinen, Karjalainen, Paatero, Jarvelainen, 2001, pp. 1-4). Objective audio testing procedures have begun to be put in practice, mainly in the areas of telephony and codecs such as MP3's. However, research conducted for this thesis has concluded there have not been any studies conducted for objective testing methods for any form of audio synthesis.

There are various forms of objective testing which have been approved by the International Telecommunication Union. For this research, the Perceptual Audio Quality Measurement (PAQM) and Noise-to-Mask Ratio (NMR) testing algorithms were chosen as the objective tests, and the Karplus-Strong Algorithm was the form of digital waveguide synthesis chosen to test.

## 5.2 Conclusions

An analysis of the results led to the conclusion that neither the PAQM nor NMR was an adequate objective test method for the Karplus-Strong algorithm form of waveguide physical modeling synthesis. While some of the samples passed the criteria or standard established earlier, neither test method was successful in matching the subjective listening tests. Only 71 percent of the samples passed with the PAQM, and only 57 percent passed using the NMR algorithm. Physical Modeling synthesis is a newer synthesis technique. It is considered to be more accurate than many other synthesis techniques, and therefore is only included in more expensive synthesizers. The sound quality of these synthesizers is too critical to accept 71 percent accuracy. This rate could possibly be accepted on lower-end keyboards using other synthesis techniques (see next section), but more work needs to be done towards finding an objective audio quality test that is accurate enough to replace subjective listening tests for Karplus-Strong waveguide synthesis.

Another interesting thought is related to the listening tests that were administered. Is there a need to establish a tone test using seven notes with the twenty participants to statistically determine homogeneity among the groups? Should pre-testing be administered in order to obtain a smaller standard deviation and error rate? These are internal changes that could possibly enhance this research. Other external test recommendations are found in the next section.

## 5.3 Recommendations for Future Research

There are many other areas and directions that can be studied in relation to the research conducted as part of this thesis. Starting with the audio synthesis portion of the research, there are other forms of waveguide synthesis that could be tested besides the Karplus-Strong algorithm which was used in this research. McIntyre, Schumacher, and Woodhouse Synthesis would be another waveguide synthesis method which could be tested using the same objective testing procedures. Also from chapter 2, there are other forms of physical modeling synthesis outside of waveguide synthesis that could be tested. Finite Element Methods, Modal Synthesis Methods, Banded Waveguides, and Combination Systems are all potential candidates for objective testing.

Another interesting research area related to this thesis would be to test other synthesis methods besides physical modeling. FM synthesis, subtractive synthesis, dynamic convolution, wavetable synthesis (Bristow-Johnson, 1996, pp. 1-26), and (Lee, Horner, 1999, pg. 101), as well as sampling are all technologies that haven't had testing done to find possible objective quality measurement techniques to replace subjective listening tests. Different levels of accuracy would need to be defined in correlation to the form of audio. For example, it may be determined that 71 percent accuracy would be sufficient for sampling technologies used on entry-level keyboards.

Other musical instruments could be tested as well. This thesis tested plucked string sounds only, however the same test could be carried out for bowed string sounds, hammered strings sounds, as well as woodwind, brass, or percussion instruments. The actual musical range could be tested differently as well. This research tested notes from the seven notes from a C major diatonic scale, beginning at 261.6 Hz and ending at 466.1 Hz. While these notes make up the most commonly used octave in traditional music, the same testing procedures used in this thesis could be tested for both an octave below, an octave above, as well as other scales.

The other research area for future work is the actual objective tests themselves. The PAQM and NMR were the two methods chosen for this work because they are currently more prominent in both academia and industry. However there have been other audio quality objective tests standardized by the International Telecommunication Union. Some of these include the Disturbance Index (DIX), Objective Audio Signal Evaluation (OASE), Perceptual Evaluation (PERCEVAL), Perceptual Objective Measurement (POM), and the toolbox approach. Research should be conducted to determine if any of these other objective audio quality measurements can successfully replace subjective listening tests when used with audio synthesis techniques. The PAQM and NMR tests are very similar to one another, and they produced very similar results. The PERCEVAL and POM tests are also similar to the PAQM and NMR tests, and their respective results could very well be similar to the results found in this research. Therefore it is recommended that specific research tests be carried out using the DIX, OASE, and toolbox approach objective test methods.

# 6 References

Ashman, M. (2002). *Physical Modeling Synthesis.* 13 March 2002, available from http://www.cs.caltech.edu/~ashman/ee107b/project107b.html; Internet; accessed 26 September 2004.

Bank, B. & Valimaki, V. (2003). *Robust Loss Filter Design for Digital Waveguide Synthesis of String Tones.* IEEE Signal Processing Letter, Vol. 10, No. 1, January 2003.

Beerends, J., Hekstra, A., Rix, A. & Hollier, M. (2001). *Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II – Psychoacoustic Model.* . for publication in the J. Audio Engineering Society, May 27, 2001.

Beerends, J. (1998). *Audio Quality Determination Based On Perceptual Measurement Techniques.* In M. Kahrs & K. Brandenberg (Eds.), Applications of Digital Signal Processing to Audio and Acoustics (pp. 1-38). Kluwer Academic Publishers.

Bilbao, S. (2004). *Wave and Scattering Methods for Numerical Simulation.* John Wiley & Sons, Ltd., 2004.

Bilbao, S. (2001). *Wave and Scattering Methods for the Numerical Integration of Partial Differential Equations.* (Doctoral dissertation, Stanford University, 2001). Available on-line at http://ccrma.stanford.edu/~bilbao/master/goodcopy.html

Bonebright, T., Miner, N., Goldsmith, T. & Caudell, T. (1998). *Data Collection and Analysis Techniques for Evaluating the Perceptual Qualities of Auditory Stimuli.* . in Proceedings of the International Conference on Auditory Display (November 1-4, 1998).

Bristow-Johnson, R. (1996). *Wavetable Synthesis 101: A Fundamental Perspective.* AES 101 Conference Preprint #4400, 1996.

Burk, P., Polansky, L., Roberts, M., Rockmore, D., Repetto, D. (2004) *Music and Computers: A Theoretical and Historical Approach.* Key College Publishing, 2004.

Cook, P. (2004) *Theory of Banded Waveguides*. Computer Music Journal, Vol. 28, Issue 1 (March 2004), pp. 37-50.

Cook, P. (2002). *Real Sound Synthesis for Interactive Applications*, A.K. Peters, L.T.D., 2002.

Cook, P. (2002). *Sound Production and Modeling.* IEEE Computer Graphics and Applications, Vol. 22, pp. 23-27.

Cook, P. (2000). *Physically-Based parametric Sound Synthesis and Control (power point title) lecture notes from the SIGGRAPH 2000 Conference.*

Cook, P. (1998). *Toward the Perfect Audio Morph? Singing Voice Synthesis and Processing.* .in Proceedings of the Workshop on Digital Audio Effects (November 19-21, 1998).

Digital Waveguide Synthesis, 6 September 2006, available from http://en.wikipedia.org/w/index.php?title=Digital_waveguide_synthesis&oldid=74042907; Internet; accessed 5 September 2006.

Erkut, C. (2002). Aspects in Analysis and Model-Based Sound Synthesis of Plucked Strings Instruments. (Doctoral Dissertation, Helsinki University of Technology, 2002).

Erkut, C. & Karjalainen, M. (2002). *Finite Difference Method vs. Digital Waveguide Method in String Instrument Modeling and Synthesis.* . in Proceedings of the International Symposium on Musical Acoustics (December 9-13, 2002).

Erkut, C. & Karjalainen, M. (2002) *Virtual Strings Based on a 1-D FDTD Waveguide Model: Stability, Losses, and Traveling Waves.* .in Proceedings of the Audio Engineering Society 22nd International Conference (June 15-17, 2002).

Erkut, C. & Valimaki, V. (2000). *Model-Based Sound Synthesis of Tanbur, a Turkish Long-Necked Lute.* . in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (June 7-9, 2000).

Esquef, P. (2004). Model-Based Analysis of Noisy Musical Recordings with Application to Audio Restoration. (Doctoral Dissertation, Helsinki University of Technology, 2004).

Essl, G. (2003). *The Displaced Bow and Aphisms: Abstract Physically Informed Synthesis Methods for Composition and Interactive Performance.* . in Proceedings of the Florida Electronic Music Festival (April 3-5, 2003).

Essl, G. (2002). Physical Wave Propagation Modeling for Real-Time Synthesis of Natural Sounds. (Doctoral Dissertation, Princeton University, 2002).

Essl, G. & Cook, P. (2000). *Measurements and Efficient Simulations of Bowed Bars.* Journal of the Acoustical Society of America, Vol. 108 No. 1, July 2000.

Fitz, K. & Haken, L. (2002). *On the Use of Time-Frequency Reassignment in Additive Sound Modeling.* Journal of Audio Engineering Society, Vol. 50, No. 11, pp. 879-893.

Hiipakka, J. (1999). Implementation and Control of a Real-Time Guitar Synthesizer. (Master's Thesis, Helsinki University of Technology, 1999).

International Telecommunication Union – Radiocommunication Group. (2001). *Method For Objective Measurements Of Perceived Audio Quality* (BS.1387-1).

International Telecommunication Union – Telecommunication Group (1996). *Methods For Subjective Determination Of Transmission Quality* (P.800).

Jarvelainen, H. (2001). *Applying Perceptual Knowledge to String Instrument Synthesis.* . in Proceedings of the Workshop on Current Research Directions in Computer Music (November 15-17, 2001).

Karjalainen, M. (2004). *Discrete-Time Modeling and Simulation of Physical (Acoustic) Systems (power point title) lecture notes from Seminar in Acoustics, Helsinki University of Technology.*

Karjalainen, M. (2004). *Discrete-Time Modeling and Synthesis of Musical Instruments.* . in Proceedings of the Joint Baltic-Nordic Acoustics Meeting (June 8-10, 2004).

Karjalainen, M. (2003). *BlockCompiler – A Research Tool for Physical Modeling and DSP.* . in Proceedings of the Sixth International Conference on Digital Audio Effects (September 8-11, 2003).

Karjalainen, M. (2002). *1-D Digital Waveguide Modeling For Improved Sound Synthesis.* . in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (May 13-17, 2002).

Karjalainen, M., Valimaki, V. & Esquef, P. (2002) *Efficient Modeling and Synthesis of Bell-Like Sounds.* . in Proceedings of the 5th International Conference on Digital Audio Effects (September 26-28, 2002).

Karjalainen, M. (1999). *New Techniques and Effects in Model-Based Sound Synthesis.* . in Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (December 9-11, 1999).

Karjalainen, M., Valimaki, V. & Tolonen, T. (1998). *Plucked String Models: From the Karplus-Strong Algorithm to Digital Waveguides and Beyond.* Computer Music Journal, vol. 22, pp. 17-32, Fall, 1998.

Karjalainen, M., Valimaki, V. & Janosy, Z. (1993). *Towards High-End Quality Sound Synthesis of the Guitar and String Instruments.* . in Proceedings of the International Computer Music Conference (September 10-15, 1993).

Laurson, M., Erkut, C., Valimaki, V., & Kuushankare, M. (2001). *Methods for Modeling Realistic Playing in Acoustic Guitar Synthesis.* Computer Music Journal, vol. 25, n0. 3, pp. 38-49, 2001.

Lee, K. & Horner, A. (1999). *Modeling Piano Tones with Group Synthesis.* Journal of Audio Engineering Society, Vol. 47, No.3, pp. 101-111.

Lehman, S. (1996). Physical Modeling Synthesis, 1996, available from http://www.harmony-central.com/Synth/Articles/Physical_Modeling/; Internet; accessed 10 October, 2004.

Masri, P. (1996). Computer Modelling of Sound for Transformation and Synthesis of Musical Signals. (Doctoral Dissertation, University of Bristol, 1996).

O'Brien, J., Cook, P. & Essl, G. (2001). *Synthesizing Sounds from Physically Based Motion.* . in Proceedings of ACM SIGGRAPH: 529-536 (2001).

Pakarinen, J. (2004). Spatially Distributed Computational Modeling of a Nonlinear Vibrating String. (Master's Thesis, Helsinki University of Technology, 2004).

Painter, T. & Spanias, A. (2000). *Perceptual Coding of Digital Audio.* . in Proceedings of the IEEE, Vol. 88, No. 4, April 2000.

Penttinen, H., Karjalainen, M., Paatero, T. & Jarvelainen, H. (2001). *New Techniques to Model Reverberant Instrument Body Responses.* . in Proceedings of the International Computer Music Conference (September 17-23, 2001).

Perceptual Audio Quality Measurement, available from http://www.opticominstruments.com/paqm.shtml; Internet; accessed 5 October, 2005.

Physical Modeling Synthesis, December, 2005, available from http://encyclopedia.thefreedictionary.com/Physical%20modeling%20synthesis; Internet; accessed 2 October, 2006.

Rabenstein, R. & Trautmann, L. (2002). *Digital Sound Synthesis of String Instruments with the Functional Transformation Method.* Signal Processing, Vol. 83, pp. 1673-1688, Aug. 2003.

Roads, C. (1996). *The Computer Music Tutorial.* The MIT Press, 1996.

Russ, M. (1994). *Yamaha VL1*. Sound on Sound, July, 1994. Available on-line at http://www.soundonsound.com/sos/1994_articles/jul94/yamahavl1.html

Savioja, L. & Valimaki, V. (1999). *Reduction of the Dispersion Error in the Interpolated Digital Waveguide Mesh Using Frequency Warping.* . in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, vol. 2 (New York), pp. 973-976, IEEE Press, March 15-19, 1999.

Singer, C. (2004). Physical Modeling on the Nord Modular G2, available from http://chet.getchwood.com/G2-Tutorial/Index.htm; Internet; accessed 5 April, 2006.

Smith, J (2006). *Physical Audio Signal Processing: Digital Waveguide Modeling of Musical Instruments and Audio Effects, June 2006 Draft.* Center for Computer Research in Music and Acoustics (CCRMA), Stanford University. Web published at http://ccrma.stanford.edu/~jos/pasp/.

Smith, J. (2004). *On the Equivalence of the Digital Waveguide and Finite Difference Time Domain Schemes.* Unpublished manuscript.

Smith, J. (1996). *Physical Modeling Synthesis Update.* Computer Music Journal, Vol. 20, No. 2, pp. 44-56, Summer 1996.

Smith, J. & Van Duyne, S. (1995). *Commuted Piano Synthesis.* . in Proceedings of the 1995 International Computer Music Conference, pp. 319-326, 1995.

Smith, J. (1992). *Physical Modeling Using Digital Waveguides.* Computer Music Journal, vol. 16, pp. 74-91, Winter 1992, Special Issue: Physical Modeling of Musical Instruments, Part I.

Sondius –XG, 13 May, 1999, available from http://www.sondius-XG.com/; Internet; accessed 5 April, 2006.

Steiglitz, K (1996). *A DSP Primer with Applications to Digital Audio and Computer Music.* Addison-Wesley Publishing Company, Inc. 1996.

Tapia, D. *An Overview of Physical Modeling Synthesis: Its Origins and Applications.* University of Northern Colorado Music Technology.

Thiede, T. & Kabot E. (1996). *A New Perceptual Quality Measure for Bit Rate Reduced Audio.* . in Proceedings of the 100th Audio Engineering Society Convention (May 11-14, 1996).

Valimaki, V. (1995). Discrete-Time Modeling of Acoustic Tubes using Fractional Delay Filters. (Doctoral Dissertation, Helsinki University of Technology, 1995).

Van den Doel, K., Pai, D., Adam, T., Kortchmar, L. & Pichora-Fuller, K. (2002). *Measurements of Perceptual Quality of Contact Sound Models.* . in Proceedings of the 2002 International Conference on Auditory Display (July 2-5, 2002).

Van den Doel, K. & Pai, D. (1996). *Synthesis of Shape Dependent Sounds with Physical Modeling.* . in Proceedings of the International Conference on Auditory Display, ICAD (1996).

Van Duyne, S. & Smith, J. (1999). *Physical Modeling with the 2-D Digital Waveguide Mesh.* . in Proceedings of the International Computer Music Conference (September 1993).

Yerrick, D. (2006) Karplus-Strong string synthesis, 22 July, 2006, available from http://en.wikipedia.org/wiki/Karplus-Strong; Internet; accessed 4 September, 2006.

# Appendices

# Appendix A. Digital Waveguide Source Code

The following source code is a C++ example of a plucked string digital waveguide synthesis using the Karplus-Strong algorithm. This is taken from the Synthesis Toolkit (STK) open source files. The comments reflect the digital waveguide technology patents held and used by Yamaha.

```
/***************************************************/
/*! \class Plucked
   \brief STK plucked string model class.

   This class implements a simple plucked string
   physical model based on the Karplus-Strong
   algorithm.

   This is a digital waveguide model, making its
   use possibly subject to patents held by
   Stanford University, Yamaha, and others.
   There exist at least two patents, assigned to
   Stanford, bearing the names of Karplus and/or
   Strong.

   By Perry R. Cook and Gary P. Scavone, 1995 – 2005.
*/
/***************************************************/

#include "Plucked.h"

Plucked :: Plucked(StkFloat lowestFrequency)
{
 length_ = (unsigned long) (Stk::sampleRate() / lowestFrequency + 1);
 loopGain_ = 0.999;
 delayLine_.setMaximumDelay( length_ );
 delayLine_.setDelay( 0.5 * length_ );
 this->clear();
```

```
}
Plucked :: ~Plucked()
{
}

void Plucked :: clear()
{
  delayLine_.clear();
  loopFilter_.clear();
  pickFilter_.clear();
}

void Plucked :: setFrequency(StkFloat frequency)
{
  StkFloat freakency = frequency;
  if ( frequency <= 0.0 ) {
    errorString_ << "Plucked::setFrequency: parameter is less than or equal to zero!";
    handleError( StkError::WARNING );
    freakency = 220.0;
  }

  // Delay = length – approximate filter delay.
  StkFloat delay = (Stk::sampleRate() / freakency) – 0.5;
  if ( delay <= 0.0 )
    delay = 0.3;
  else if ( delay > length_ )
    delay = length_;
  delayLine_.setDelay( delay );

  loopGain_ = 0.995 + (freakency * 0.000005);
  if ( loopGain_ >= 1.0 ) loopGain_ = 0.99999;
}

void Plucked :: pluck(StkFloat amplitude)
{
  StkFloat gain = amplitude;
  if ( gain > 1.0 ) {
    errorString_ << "Plucked::pluck: amplitude is greater than 1.0 … setting to 1.0!";
    handleError( StkError::WARNING );
    gain = 1.0;
  }
  else if ( gain < 0.0 ) {
    errorString_ << "Plucked::pluck: amplitude is < 0.0  … setting to 0.0!";
    handleError( StkError::WARNING );
    gain = 0.0;
  }
```

```
  pickFilter_.setPole( 0.999 – (gain * 0.15) );
  pickFilter_.setGain( gain * 0.5 );
  for (unsigned long i=0; i<length_; i++)
    // Fill delay with noise additively with current contents.
    delayLine_.tick( 0.6 * delayLine_.lastOut() + pickFilter_.tick( noise_.tick() ) );
}

void Plucked :: noteOn(StkFloat frequency, StkFloat amplitude)
{
  this->setFrequency( frequency );
  this->pluck( amplitude );

#if defined(_STK_DEBUG_)
  errorString_ << "Plucked::NoteOn: frequency = " << frequency << ", amplitude = " <<
amplitude << ".";
  handleError( StkError::DEBUG_WARNING );
#endif
}

void Plucked :: noteOff(StkFloat amplitude)
{
  loopGain_ = 1.0 – amplitude;
  if ( loopGain_ < 0.0 ) {
    errorString_ << "Plucked::noteOff: amplitude is greater than 1.0 … setting to 1.0!";
    handleError( StkError::WARNING );
    loopGain_ = 0.0;
  }
  else if ( loopGain_ > 1.0 ) {
    errorString_ << "Plucked::noteOff: amplitude is < 0.0  … setting to 0.0!";
    handleError( StkError::WARNING );
    loopGain_ = (StkFloat) 0.99999;
  }

#if defined(_STK_DEBUG_)
  errorString_ << "Plucked::NoteOff: amplitude = " << amplitude << ".";
  handleError( StkError::DEBUG_WARNING );
#endif
}

StkFloat Plucked :: computeSample()
{
  // Here's the whole inner loop of the instrument!!
  lastOutput_ = delayLine_.tick( loopFilter_.tick( delayLine_.lastOut() * loopGain_ ) );
  lastOutput_ *= 3.0;
  return lastOutput_;
}
```

# Appendix B. PAQM Source Code

The following source code is a Matlab example of the Perceptual Audio Quality Measurement (PAQM) test carried out for this research.

```
%PAQM test

%FFT – softsynth
PAQMSreload=1;
if PAQMSreload==1
   PAQMSY=wavread('Softsynth.wav');
      PAQMSyf=fft(PAQMSY(:,1));
end

PAQMSyfm=abs(PAQMSyf);

PAQMSN = size(PAQMSyfm,1);

PAQMSyfm2=PAQMSyfm(1:PAQMSN/2)';


PAQMStopFreq=44100;

PAQMSbinInc = PAQMStopFreq/PAQMSN;

PAQMSfKHz = (1:PAQMSN/2)*PAQMSbinInc/1000;  % 1000 cuz of KHz


%FFT – acoustic
PAQMAreload=1;
if PAQMAreload==1
   PAQMAY=wavread('Acoustic.wav');
      PAQMAyf=fft(PAQMAY(:,1));
end

PAQMAyfm=abs(PAQMAyf);
PAQMAN = size(PAQMAyfm,1);
```

```
PAQMAyfm2=PAQMAyfm(1:PAQMAN/2)’;
PAQMAtopFreq=44100;

PAQMAbinInc = PAQMAtopFreq/PAQMAN;

PAQMAfKHz = (1:PAQMAN/2)*PAQMAbinInc/1000;  % 1000 cuz of KHz

%softsynth
%weighting with transfer functions of the human ear
% apply outer and inner ear transfer function
% (from Terhardt), Calculating virtual pitch.  Hearing Research 1:155-18x,
% via traistan jehan
PAQMSAdB =  -3.64 * PAQMSfKHz.^(-0.8) + 6.5 .* exp (-0.6 * (PAQMSfKHz -
3.3).^2) – 10^-3 .* PAQMSfKHz.^4;
PAQMSscale = 10.^(PAQMSAdB./20);


% scaled magnitude
PAQMSsMag = PAQMSyfm2 .* PAQMSscale;



%acoustic
% apply outer and inner ear transfer function
% (from Terhardt), Calculating virtual pitch.  Hearing Research 1:155-18x,
% via traistan jehan
PAQMAAdB =  -3.64 * PAQMAfKHz.^(-0.8) + 6.5 .* exp (-0.6 * (PAQMAfKHz -
3.3).^2) – 10^-3 .* PAQMAfKHz.^4;
PAQMAscale = 10.^(PAQMAAdB./20);

% scaled magnitude
PAQMAsMag = PAQMAyfm2 .* PAQMAscale;



%softsynth
%scale    transformation from frequency to pitch
PAQMSb=26.81*PAQMSsMag./(1960+PAQMSsMag)-0.53;
PAQMSb2=PAQMSb+0.15*(2-PAQMSb).*(PAQMSb<2)+0.22*(PAQMSb-
20.1).*(PAQMSb>20.1);


%acoustic
PAQMAb=26.81*PAQMAsMag./(1960+PAQMAsMag)-0.53;
PAQMAb2=PAQMAb+0.15*(2-PAQMAb).*(PAQMAb<2)+0.22*(PAQMAb-
20.1).*(PAQMAb>20.1);
```

```
%time-frequency spreading – softsynth
PAQMSyfm2t = PAQMSyfm2';
PAQMSw3 = conv(PAQMSb2,PAQMSsMag)


%time-frequency spreading – acoustic
PAQMAyfm2t = PAQMAyfm2';
PAQMAw3 = conv(PAQMAb2,PAQMAsMag)


%comparison
PAQMtest=PAQMSw3-PAQMAw3;
PAQMfinal=abs(PAQMtest);
C = max(PAQMfinal);
D = min(PAQMfinal);
n = numel(PAQMfinal);
var1 = C + D;
var2 = var1 / n;
var3 = var2 * 100;
finalvar = fix(var3);
```

# Appendix C. NMR Source Code

The following source code is Matlab example of the Noise-To-Mask Ratio (NMR) test used in this research.

```
%NMR test

%FFT - softsynth
NMRSreload=1;
if NMRSreload==1
   NMRSY=wavread('Softsynth.wav');
      NMRSyf=fft(NMRSY(:,1));
end

NMRSyfm=abs(NMRSyf);

NMRSN = size(NMRSyfm,1);

NMRSyfm2=NMRSyfm(1:NMRSN/2)';

NMRStopFreq=44100;

NMRSbinInc = NMRStopFreq/NMRSN;

NMRSfKHz = (1:NMRSN/2)*NMRSbinInc/1000;  % 1000 cuz of KHz


%FFT - acoustic
NMRAreload=1;
if NMRAreload==1
   NMRAY=wavread('Acoustic.wav');
      NMRAyf=fft(NMRAY(:,1));
end

NMRAyfm=abs(NMRAyf);

NMRAN = size(NMRAyfm,1);
```

```
NMRAyfm2=NMRAyfm(1:NMRAN/2)';
NMRAtopFreq=44100;
NMRAbinInc = NMRAtopFreq/NMRAN;
NMRAfKHz = (1:NMRAN/2)*NMRAbinInc/1000;  % 1000 cuz of KHz

%softsynth
%weighting with transfer functions of the human ear
% apply outer and inner ear transfer function
% (from Terhardt), Calculating virtual pitch.  Hearing Research 1:155-18x,
% via traistan jehan
NMRSAdB =  -3.64 * NMRSfKHz.^(-0.8) + 6.5 .* exp (-0.6 * (NMRSfKHz -3.3).^2) -
10^-3 .* NMRSfKHz.^4;
NMRSscale = 10.^(NMRSAdB./20);


% scaled magnitude
NMRSsMag = NMRSyfm2 .* NMRSscale;



%acoustic
% apply outer and inner ear transfer function
% (from Terhardt), Calculating virtual pitch.  Hearing Research 1:155-18x,
% via traistan jehan
NMRAAdB =  -3.64 * NMRAfKHz.^(-0.8) + 6.5 .* exp (-0.6 * (NMRAfKHz -3.3).^2) -
10^-3 .* NMRAfKHz.^4;
NMRAscale = 10.^(NMRAAdB./20);


% scaled magnitude
NMRAsMag = NMRAyfm2 .* NMRAscale;



%softsynth
%scale transformation from frequency to pitch
NMRSb=26.81*NMRSsMag./(1960+NMRSsMag)-0.53;
NMRSb2=NMRSb+0.15*(2-NMRSb).*(NMRSb<2)+0.22*(NMRSb-
20.1).*(NMRSb>20.1);

%acoustic
NMRAb=26.81*NMRAsMag./(1960+NMRAsMag)-0.53;
NMRAb2=NMRAb+0.15*(2-NMRAb).*(NMRAb<2)+0.22*(NMRAb-
20.1).*(NMRAb>20.1);


%time-frequency spreading - acoustic
```

```
NMRAyfm2t = NMRAyfm2';
NMRAw3 = conv(NMRAb2,NMRAsMag)


%comparison
BS = sum(NMRNw3);
BA = sum(NMRSb2);
abBS = abs(BS);
abBA = abs(BA);
var1B = abBS / abBA;
el = numel(NMRAw3);
var2B = var1B / el;
var3B = var2B * 10;
finalvarB = round(var3B);
```

# Appendix D. Listening Test Survey Sheet

For each of the following samples, give a score from the following guidelines based on the similarity of the sounds:
5 – Sound Nothing Alike
4 – Doesn't Sound Very Similar
3 – Sound Somewhat Similar
2 – Sound Very Similar
1 – Sound Exactly Alike

**Sample A -**
**Sample B -**
**Sample C -**
**Sample D -**
**Sample E -**
**Sample F -**
**Sample G -**

Describe the audio system used to listen to the samples.

Personal Background Questions:

What is your primary musical area of expertise? (vocal, guitar/bass, keyboard/piano, percussion, DJ, woodwind, brass, orchestral strings, recording, etc.)

Have you ever been enrolled in a music class?

If yes, what was/is the highest level? (college, high school, private instruction, etc.)

Thank You!