Audio Engineering Society

Convention Paper 6525

Presented at the 119th Convention
2005 October 7–10 New York, NY, USA

Perceptual Modeling of Piano Tones

Brahim Hamadicharef[1], and Emmanuel Ifeachor[1]

[1] *University of Plymouth, Plymouth, PL4 8AA, Devon, U.K.*

Correspondence should be addressed to Brahim Hamadicharef (bhamadicharef@plymouth.ac.uk)

ABSTRACT

A modeling system for piano tones is presented. It fully automates the modeling process and includes the following three mains stages: sound analysis, sound synthesis and sound quality assessment. High quality piano sounds are analysed in time and frequency domain. Analysis results are then used to design filter models matching the string resonance and create excitation signals using an inverse filtering technique for the excitation-filter synthesis model. The impact of each sound model parameter onto the perceived sound quality has been assessed using Perceptual Evaluation of Audio Quality (PEAQ) algorithm. This is helping to optimise the DSP resources requirements for real-time implementation onto multimedia PC and FPGA-based hardware.

1. INTRODUCTION

This paper presents a new concept of perceptual modeling of piano tones. This modeling system implements a new framework for perceptual sound modeling. This work is part of an on-going research on the potential of perceptual audio quality based modeling of musical instruments.

The modeling system, presented in this paper, is based on the multi-channel excitation/filter sound synthesis technique described by Laroche and Meillier [1]. This synthesis technique aims to model piano tones from a *digital signal processing* point of view. This can be easily understood by making the analogy with a physical model of a piano string. In simple terms, excitation signal aims to model the impact of the hammer on the string, while the piano string is modeled using digital filters which corresponds to the resonance properties of piano string.

The modeling framework uses the Perceptual Evaluation of Audio Quality (PEAQ) algorithm [2] to objectively assess the final quality of the synthetic sound. It facilitates the automation of the sound modeling process, also allowing to evaluate the impact of each sound synthesis parameter on the final sound quality. This framework for perceptual sound modeling is generic and can potentially be used to model other instruments using other types of sound synthesis.

The paper is organised as follow. In Section 2, the concept of perceptual modeling of piano tones is presented with a description of the key elements of the modeling system. In Section 3, we present results from a series of experiments conducted with the modeling system. Finally, Section 4 concludes the paper.

2. PERCEPTUAL MODELING OF PIANO TONES

2.1. System Overview

A conceptual diagram of the perceptual modeling framework for piano tones is shown in Figure 1. Original piano sounds are loaded and the sound analysis used to extract time-frequency sound features. These features are then used to generate parameters (excitation signal and digital filter) for the sound synthesis to recreate a synthetic version of the original sound. The sound quality assessment stage uses both original and synthetic sounds to calculate a sound quality index. This index is used to evaluate the sound modeling process. The system is implemented in MATLAB (Release 7.0.4) runs on a PC workstation and the the PEAQ algorithm from the OPERA *Voice/Audio Quality Analyzer* (Opticom GmbH, Germany).

High quality piano sounds were provided by our industrial collaborators. Few recordings from sample CDs and various commercially available piano synthesizers were also used. Sounds were all 16-bit in WAVE file format, sampled at 48 kHz, with stereo recordings converted to mono.

2.2. Sound Analysis

The sound analysis is based on a technique combining bandpass filters to isolate each harmonic and applying Teager Energy Operator (TEO) to extract both amplitude and frequency behaviour of each of the sound components [3]. As shown in Figure 2, we can extract easily most of the main harmonics of the piano sound as they exhibit large peaks in spectrum. The accuracy is important as it might affect the amplitude and frequency of harmonics in the final sound, thus the frequency result of FFT analysis is used, for there are sometimes slight differences between the theoretical values and practical ones. Figure 3 shows the amplitude envelopes obtained by TEO energy separation method. These envelopes are used in to design the digital string filters.

2.3. Excitation Signal

The excitation signal emulates the action of the hammer on the piano string. As described in [1], excitation signals can be obtain from time domain technique using an *inverse filtering* method or frequency domain method such as a least-squares deconvolution, the common excitation defined as the *inverse FFT* of:

$$E(f) = \frac{\sum_{n=1}^{N} H_i^*(f) S_i(f)}{\sum_{n=1}^{N} H_i^*(f) H_i(f)} \quad (1)$$

where $H_i(f)$ is the complex transfer function of filter H_i, and $S_i(f)$ is the Fourier transform of input signal s_n.

As shown in Figure 4, the excitation signal can be considered as short burst of noise with a decaying amplitude envelope. As excitation signals need to be stored in memory (in a typical electronic musical instrument), the task becomes either to model them or use techniques to reduce their memory storage requirements. It is obvious that reducing the size of these signals will greatly influence the sound quality of the re-created piano tones. In this work, we are particularly interested in assessing the perceptual impact onto the synthetic sound quality of such data reduction. Laroche and Meillier proposed various methods to reduce the complexity of their sound model. As shown Figure 5, one method is to create a common excitation signal that could be used for the synthesis of individual sound.

2.4. String Filter Design

The piano strings are modeled using digital filters, based on second-order cosine sections [4]. Figure 6 shows an example of such complex filter example for a key E5 piano sound. The filter models the resonances of the piano string and needs ideal to perfectly match the harmonics of the sound. These peaks (as shown in Figure 2) intuitively relate to the position of the poles of the digital filter. They are obtained by tracking both frequency and amplitude envelope of the harmonics resulting from the sound analysis [5]. The stability of the digital filter must be ensured, and thus all poles located within the unit

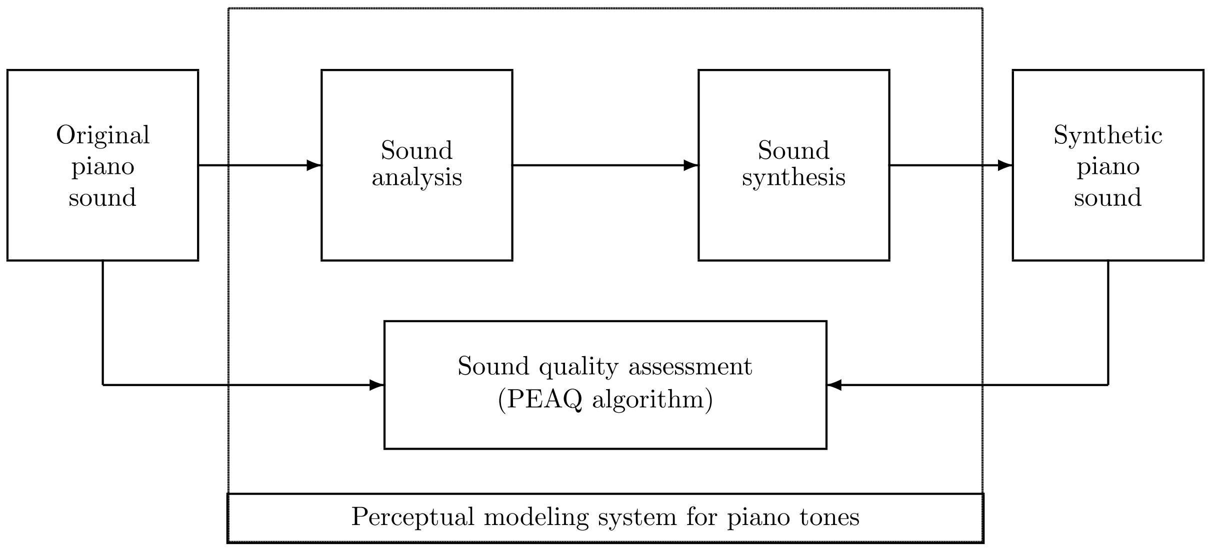

Fig. 1: Conceptual diagram of perceptual modeling framework

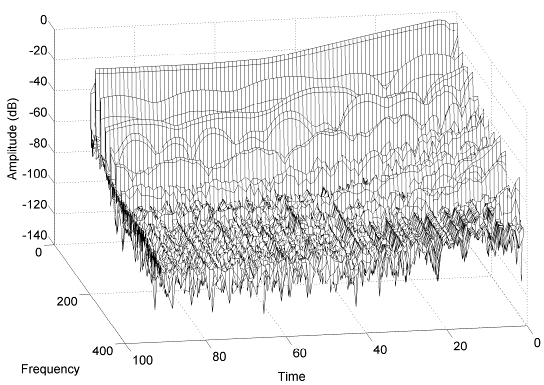

Fig. 2: Power spectrum of piano tone (key E5)

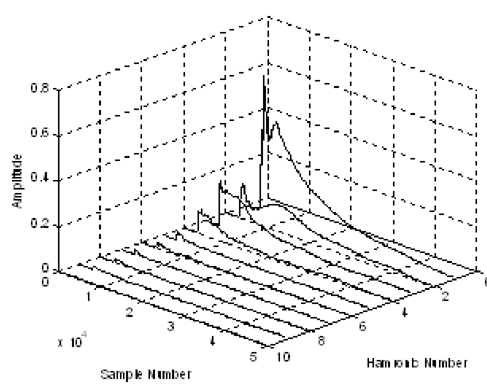

Fig. 3: Amplitude results of TEO

circle. One important issue in digital filter design is the quantisation effects on the filter coefficients and is especially important when the implementation is carried out on a fixed-point DSP hardware for example. Thus, there is a need to evaluate the impact of filter quantisation onto the final quality of piano tones.

2.5. Sound Quality Assessment

The sound quality assessment uses the Perceptual Evaluation of Audio Quality (PEAQ) algorithm [2] based on the ITU-R BS.1387 [6]. As shown in Figure 7, PEAQ consists of a perceptual analysis stage, a feature extraction and a cognitive model. The perceptual analysis replicates the human hearing system. Perceptual analysis results from the reference

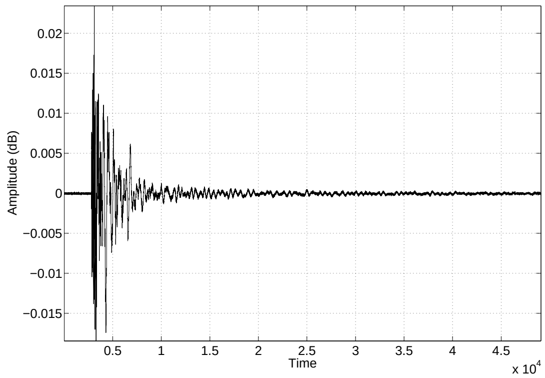

Fig. 4: Example of excitation signal

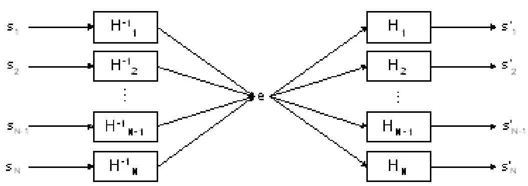

Fig. 5: Synthesis system of common excitation

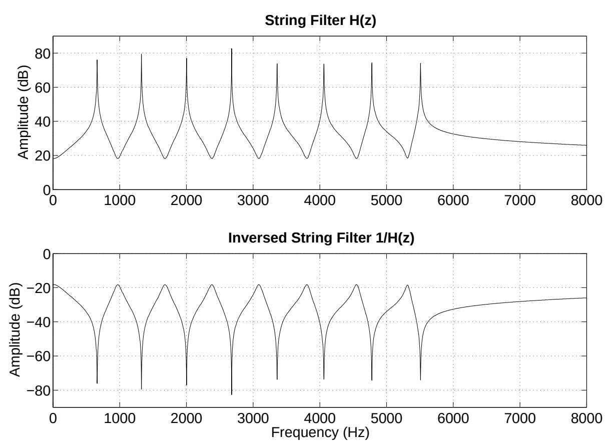

Fig. 6: Magnitude transfer function of string filter, key E5

and test sounds are used by the feature extraction to produce Model Variables (MOVs) which quantify sound features such as modulation, noise and loudness. The cognitive model, model human judgement, combines these MOVs to forms the Objective Difference Grade (ODG). This single quality index ranges from 0 (labeled as *imperceptible* audio degradation)

ODG value	Meaning
0.0	Imperceptible
-1.0	Perceptible but not annoying
-2.0	Slightly annoying
-3.0	Annoying
-4.0	Very annoying

Table 1: Objective Difference Grade (ODG) and their meaning

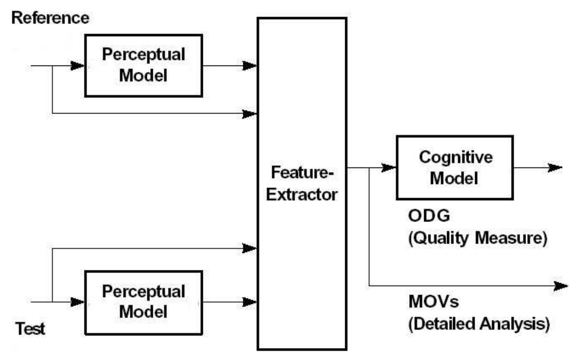

Fig. 7: Model for Perceptual Evaluation of Audio Quality (PEAQ)

down to -4 (labeled as *very annoying*), with value of -1 corresponding to the threshold for a *perceptible* degradation. Table 1 gives a summary of ODG values and their respective meaning. Detailed description of the PEAQ algorithm can be found in in [6].

3. EXPERIMENTS AND RESULTS

Experiments have been conducted applied to the excitation signal and digital filters. Excitation signals data reduction mechanisms included shortening the excitation signal by applying different windowing functions, grouping excitation signals into a composite excitation signal, use of interpolation / extrapolation techniques, etc. All these were for cases of individual sound, sounds within same octave and various key combinations. Filter models reduction were also carried out with filter type and filter order, but primarily focusing on quantisation effects on the filter coefficients.

3.1. Reduction of the excitation signal

Three type of data reduction of the excitation signal have been investigated and their impact on the perceived sound quality assessed using PEAQ. These data reduction include a process of windowing the excitation (shortening the length of the excitation signal), use of a common excitation signal from close keys and finally the use of interpolating technique to recreate excitation signals from existing excitation signals.

3.1.1. Windowing the excitation

Various type of window functions have been used to evaluate their effect. These windows are commonly used in signal processing. They include triangle, rectangular, mixed, logarithm and exponential. A rectangular window will shorten the excitation signal while other windows will add a progressive smoothing effect the excitation overall envelope. ODG scores are summarised in Table 2. Results are showing that windowing the excitation signal can degrades dramatically the final sound quality (ODG scores between -3 and -4 stand for *very annoying*). Further listening tests with the help of an audio expert also helped to confirm this sound degradation and modification of the final timbre of the piano tones. Shortening the excitation signal tend to remove components which control the resonant and beating aspect of piano sounds, resulting in some very unnatural steadiness of the sustain part.

3.1.2. Common excitation

En example of common excitation generated from 3 keys is presented in Figure 8. Table 3 is showing this error for 3 keys ($C5$, $C5\sharp$ and $D5$) and Table 4 for an example with 5 keys ($C5$, $C5\sharp$, $D5$, $D5\sharp$ and $E5$). Mean Squared Error (MSE) values are representing difference between the estimated excitation and the original excitation. Results are showing that as we increase the number of sounds used in the creation of the common excitation signal, the final synthetic sound quality drops (i.e. lower ODG values and higher MSE).

3.1.3. Interpolation of excitation

We also investigated interpolation of the excitation signals within key intervals. Combinations tested are: COMB1) white keys to obtain black keys (e.g. use the excitations of keys F5 and G5 to generate the excitation for $F5\sharp$), COMB2) sound in a chord (e.g. D5 and $F5\sharp$ to model A5), COMB3) same tone of different octaves (e.g. C4 and C6 to model C5), COMB4) adjacent keys (e.g. C5 and D5 to model E5). Furthermore, we were interested in assessing the possibility of using common excitation signals for multiple keys.

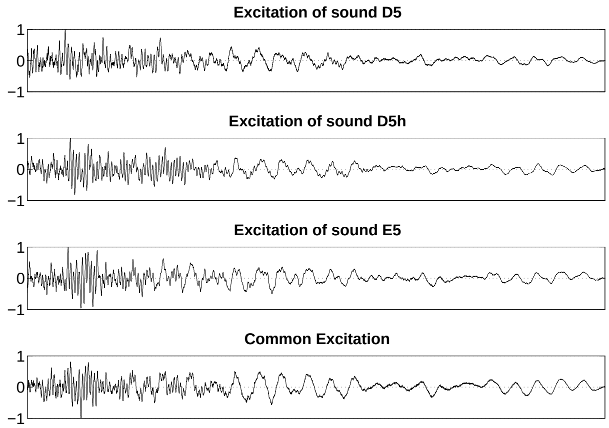

Fig. 8: Common excitation

Key	Mean ODG	MSE of Excitation
$C5$	-1.7155	0.0012421
$C5\sharp$	-1.4428	0.0015984
$D5$	-1.6626	0.0014435

Table 3: ODG results and excitation MSE of excitation vs. common excitation for 3 keys

Key	Mean ODG	MSE of Excitation
$C5$	-2.3423	0.0023425
$C5\sharp$	-1.8432	0.0017373
$D5$	-1.9461	0.0026022
$D5\sharp$	-2.2632	0.0017150
$E5$	-2.1640	0.0023920

Table 4: ODG results and excitation MSE of excitation vs. common excitation for 5 keys

Window Length	Triangle	Rectangular	Mixed	Logarithm	Exponential
4K	-3.557	-3.653	-3.559	-3.568	-3.532
8K	-3.590	-3.374	-3.422	-3.505	-3.569
16K	-3.495	-3.007	-3.149	-3.284	-3.589
32K	-3.237	-2.120	-2.594	-2.906	-3.654
64K	-2.770	-0.682	-1.446	-2.041	-3.621

Table 2: ODG results of windowing the excitation

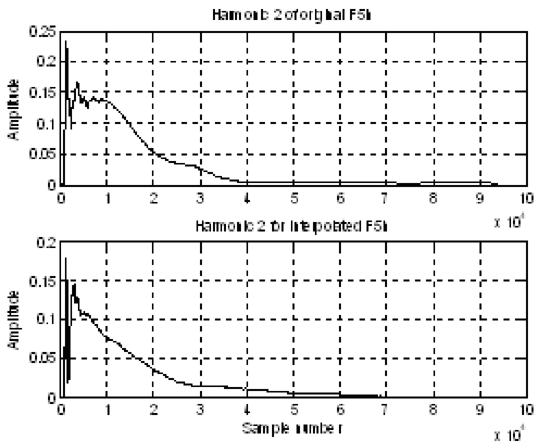

Fig. 9: Comparing 2^{nd} harmonic in COMB1

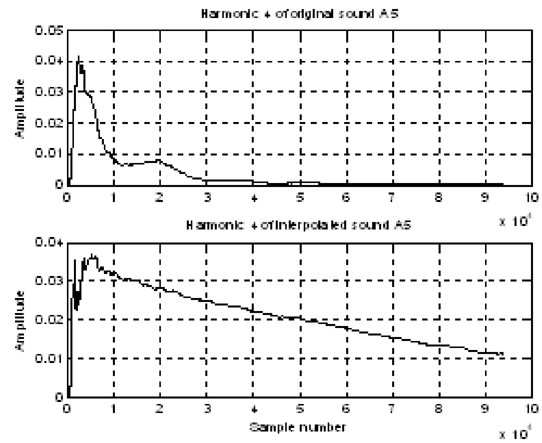

Fig. 10: Comparing 4^{th} harmonic of COMB2

Result quality ODG scores of COMB1): -3.671; COMB2): -3.7362; COMB3): -3.7138; COMB4): -3.8637; using common excitation: -3.7501. All these results show a *very annoying* sound quality and demonstrate bad choice. Figure 9 and Figure 10 show the effects of the interpolation process, changing the decay rates of the resulting individual harmonics. Further studies, comparing the spectrum of estimated and unknown excitations signals, indicated that the estimated excitation contains the frequency components from all known excitations, but lack of some the components of the modeled excitation (which are known from the analysis of the excitation signal of this specific key).

3.2. Filter Coefficients Quantisation

To assess the perceptual impact of such effects, we have carried out quantisation on the filter coefficients. Filter coefficients quantisation changes the filter coefficients values and thus directly the filter frequency response, having direct effect on the final sound quality of the piano tones.

Figure 11 and Figure 12 show the ODG versus the quantisation (defined as word length in bits) for a key E5 and a key C4. It is clear from these figures that the quantisation of the filter coefficients leads to dramatic quality drops. Listening tests also pointed out some ringing audible effect. Looking at the threshold of ODG value -1, key E5 only requires (at least) 34-bits while key C4 42-bits precision. This is also showing that lower pitched sounds requires higher accuracy to their filter model parameters than higher pitched sounds.

4. CONCLUSIONS

In this paper, a fully automated modeling system for piano tones is presented. This is part of an on-going

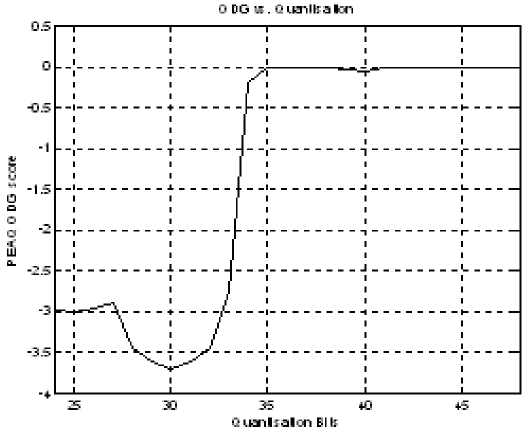

Fig. 11: ODG versus quantisation (E5 key)

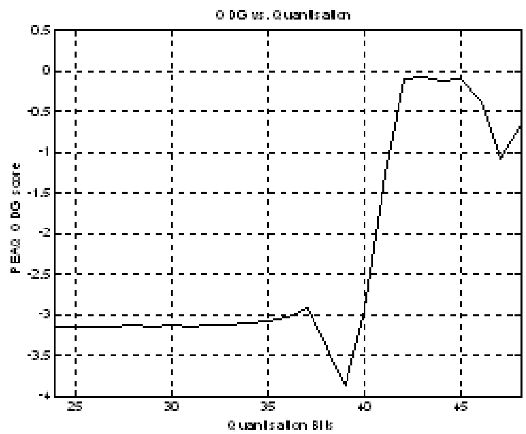

Fig. 12: ODG versus Quantisation (C4 key)

research on the potential of perceptual audio quality based modeling of musical instruments. The modeling system for piano tones consists of three stages: sound analysis, sound synthesis and objective sound quality prediction.

High quality piano sounds provided by our industrial collaborators have been analysed in time and frequency domain and the results used to design filter models and create excitation signals. This follows the modeling concept for excitation-filter sound synthesis. The filter models try to match as closely as possible the resonance of the piano string. The excitation signals are obtained using an inverse filtering technique. The signals correspond to the percussion phenomena of the hammer on the piano string. The best synthesis parameters are used to create a reference synthetic sound.

A series of experiments were carried out to evaluate the impact of the sound model parameters on the final perceived sound quality using the Perceptual Evaluation of Audio Quality (PEAQ) algorithm. This was motivated by the need to find out 1) which aspect of the sound synthesis model has the most perceptual influence on the final synthetic sound and 2) the degree to which synthesis parameters can reduced before degradation in the quality of the synthetic sound becomes perceptible but not annoying (corresponding to an Objective Difference Grade (ODG) value of -1 by the PEAQ algorithm).

Excitation signals data reduction mechanisms we investigated by shortening the excitation signal while applying different windowing functions, grouping excitation signals into one composite signal, using of interpolation / extrapolation techniques. All these were for cases of individual sound, sounds within same octave and various keys combinations. Models parameter reduction for the filter included the study of quantisation effects on the filter coefficients and different filter types and filter orders.

Results have shown that both reduction of excitation signals and filter quantisation can degrade dramatically the final sound quality, with the excitation signal having greatest impact, especially on the final timbre of the piano tones. This is of great importance as from a DSP implementation point of view, the memory space used for filter coefficients is much smaller than for the excitation signals.

The modeling system is able to model piano tones with little perceptual distortion. It has been found that more work is needed on lower pitch sounds as they are more difficult modeling task than high pitched ones. Our collaborating audio expert noted the quality of the system good/high in listening tests.

This work is helping to optimise the DSP resources requirements for real-time sound synthesis implementation onto multimedia PC, DSP and FPGA-based hardware and find application as sound generator of electronic piano.

5. ACKNOWLEDGEMENTS

This research is supported by the Engineering and Physical Science Research Council (EPSRC) under the grant (GR/S00859). We also acknowledge support from the UK Government Science Research Investment Fund (SRIF) initiative for the purchase of the Opera *Voice/Audio Quality Analyzer* (Opticom GmbH). The authors would like to thank to Dr Rob Clark (Allen & Heath) for his advice on digital filter design and Tony Koorlander (Musicom Ltd) for the listening tests.

6. REFERENCES

[1] J. Laroche and J.-L. Meillier, "Multichannel Excitation/Filter Modeling of Percussive Sounds with Application to the Piano," *IEEE Transactions on Speech and Audio Processing*, vol. 2, pp. 329–344, April 1994.

[2] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality," *Journal of the Audio Engineering Society*, vol. 48, pp. 3–29, January/February 2000.

[3] M. Kahrs, "Audio Applications of the Teager Energy Operator," *Proceedings of the 111th Audio Engineering Society Convention, New York, USA, Nov. 30 - Dec. 3, 2001*, 2001. Preprint 5473.

[4] E. C. Ifeachor and B. W. Jervis, *Digital Signal Processing: A Practical Approach*. Prentice Hall, second ed., 2002. ISBN 0-201-59619-9.

[5] M. Kahrs and R. Sussman, "Analysis of Musical Instrument Sounds Using Discrete Energy Separation Algorithms," *Proceedings of the 101st Audio Engineering Society Convention, Los Angeles, CA, USA, November 8-11, 1996*, 1996. Preprint 4390.

[6] ITU-R Recommendation BS.1387, "Method for Objective Measurements of Perceived Audio Quality (PEAQ)," International Telecommunication Union, Geneva, Switzerland (1998).