Audio Engineering Society

Convention Paper 5958

Presented at the 115th Convention
2003 October 10–13 New York, New York

Objective Prediction of Sound Synthesis Quality

Brahim Hamadicharef[1] and Emmanuel Ifeachor[1]

[1]Department of Communication and Electronic Engineering, University of Plymouth, Drake Circus, Plymouth Devon, PL4 8AA, UK
Correspondence should be addressed to Brahim Hamadicharef (bhamadicharef@plymouth.ac.uk)

ABSTRACT

This study is concerned with objective prediction of perceived audio quality for an intelligent audio system for modeling musical instruments. The study is part of a project to develop an automated tool for sound design. Objective prediction of subjective audio quality ratings by audio experts is an important part of the system. Sound quality is assessed using PEAQ (Perceptual Evaluation of Audio Quality) algorithm, and this greatly reduces the time-consuming efforts involved in listening tests. Tests carried out using a large database of pipe organ sounds, show that the method can be used to quantify the quality of synthesized sounds. This approach provides a basis for the development of a new index for benchmarking sound synthesis techniques.

1. INTRODUCTION

An important process in the development of sound synthesis systems, such as electronic musical instruments, is the assessment of the final perceived sound quality. Subjective listening tests with human subjects (audio experts) are commonly used to obtain accurate assessment of the final perceived sound quality. However, these tests are expensive, time consuming, require specialized sound facilities and need a large number of subjects to obtain the required accuracy. These problems have resulted in extensive research into objective audio quality metrics, i.e. computational methods that correlate well with human opinion. Increased knowledge and understanding of the complex human auditory system has recently resulted in objective quality metrics based on models of human perception [1]. Results have been promising, but much work still remains to be done before these metrics are widely adopted by the audio industry.

In this paper, objective prediction of perceived sound quality for an intelligent audio system is presented. This research project has been carried out in close collaboration with two audio companies, one of which has extensive expertise in sound synthesis of electronic pipe organs. To our knowledge, this is a unique attempt in computer music to capture and exploit, explicitly knowledge from audio experts for sound design and objective prediction of perceived sound quality.

The aim of this study is to undertake investigations into novel methods of assessing sound synthesis quality. This will form the basis of future developments of a new quality index to accurately and objectively predict sound quality for the benchmark of sound synthesis techniques.

The remainder of the paper is organized as follows. In Section 2, perceptual-based optimization of sound synthesis is briefly described. Section 3 describes the analysis of sound synthesis quality. Results are presented in Section 4 with pipe organ sound examples. Finally, Section 5 concludes this paper.

2. PERCEPTUAL-BASED OPTIMIZATION OF SOUND SYNTHESIS

The perceptual-based sound optimization of sound synthesis system, shown in Fig. 1, is part of the intelligent audio system described in [2]. It consists of four parts: a Sound Analysis Engine, a Knowledge-based Audio Feature Processing Engine, a Sound Synthesis Engine, and finally a Perceptual Error Analysis Engine. The system is used to automatically analyze acoustic recordings, extract salient sound features and process them to generate optimal sound synthesis parameters, mimicking human audio experts in the complex and time-consuming task of high quality sound design and modeling of musical instruments. We aim to fully automate the task of high quality sound design, based on the knowledge and experience of audio professionals, and to use an objective prediction of the subjective rating from listening tests by audio experts to assess the final perceived quality of sound synthesis techniques. The perceptual-based sound optimization of sound synthesis system has been implemented as a research software tool; a typical screenshot of the software tool is shown in Fig. 2.

The Sound Analysis Engine (block 1 in Fig. 1) is based on a phase vocoder analysis engine. It extracts the time varying evolution of the sound harmonic components both in amplitude and frequency. It also automatically estimates the Attack / Decay / Sustain / Release (ADSR) envelope segments. The Audio Feature Processing Engine (block 2 in Fig. 1) is our novel modeling method based on a fuzzy expert system developed in collaboration with two audio experts [2]. The fuzzy expert system emulates the decision making process of the human audio expert to generate a set of optimized sound synthesis parameters.

The Sound Synthesis Engine (block 3 in Fig. 1) is based on multiple wavetable sound synthesis with advanced modulation. The system generates sound files (standard wave files) that can be edited directly from the computer using monitor speakers or headphones, and also generates configuration files for our collaborator's electronic pipe organ musical hardware. The Perceptual Error Analysis Engine is the final part of the system (show as block 4 in Fig. 1) and is based on the PEAQ algorithm. It is used to control and optimize the knowledge-based audio features processing..

3. ANALYSIS OF SOUND SYNTHESIS QUALITY

Listening tests, which are the preferred way to assess perceived audio quality of sound synthesis, are known to be subjective, difficult to perform, time-consuming, expensive, and inconsistent. The ITU-R BS.1116 standard [4] gives guideline to perform these listening tests.

To predict the perceived quality of the sounds in an objective and reproducible manner the perceived sound quality engine exploits the Perceptual Evaluation of Audio Quality (PEAQ) algorithm [5] detailed in the ITU-R BS.1387 [6].

The original sound is used as the reference input signal and the test input signal is the synthetic sound.

PEAQ consists of a perceptual model, a feature extractor and a cognitive model. The perceptual model emulates the human hearing system, while the cognitive model reproduces the judgment made by human on the sound quality. Fig. 3 shows this generic perceptual measurement algorithm.

The outputs of the PEAQ algorithm, which includes model variables and measures of sound perception, are useful for characterizing sound synthesis artifacts as well as obtaining the final measure of sound quality (known as Objective Difference Grade (ODG) see Table 1). The Objective Difference Grade (ODG) is the output variable from the objective measure method, it ranges from 0 to –4, where 0 corresponds to imperceptible and –4 judged as "very annoying". In this work, the degradation corresponds to the difference in quality between the original sound (reference) and the synthetic sound (test) produced by the intelligent audio system.

In all our experiments, we have used the basic model of PEAQ algorithm a database of pipe organ sounds from our industrial collaborators. All the sound files are sampled at 48 kHz in 16-bit PCM. We have used the Opera "Voice/Audio Quality Analyser" from Opticom GmbH [7] and a modified version of EaQual [8]). The modified EaQual generates Matlab scripts to plot the 11 Model Output Variables (as shown in Table 2.) as well as the ODG and Distortion Index (DI).

4. RESULTS

Extensive tests were conducted using a large database of pipe organ sounds provided by one of our collaborative companies. The database included a variety of pipe organ sounds recorded in churches and cathedrals across Europe and United States of America. Tests have been performed on individual pipe organ sounds, complete manuals (Great, Swell and Choir) and whole instruments. Final sound synthesis parameters were converted and loaded into an electronic pipe organ hardware provided by the collaborating company.

The perceived sound synthesis quality was assessed by audio experts rating during listening tests and then using Perceptual Error Analysis with the basic version of the PEAQ algorithm. The listening tests have been performed at our industrial collaborator facilities, which involved assessment of single note sound as well as extended piece of music.

To illustrate our work, we have chosen typical examples of pipe organ sounds from a sound database of a reference pipe organ in Texas, USA.

4.1. Choir sound

The first example is a sound from the choir manual. The sound analysis engine estimates the fundamental to be 108.352 Hz. This sound has very clean timbre. The sound analysis resulted in 222 harmonics. Some harmonics have similar time-varying evolutions and seem very dominants. The time domain waveform of this sound is shown Fig. 4. Fig. 5 shows the sound spectrum while Fig. 6 shows the harmonics analysis. In the harmonics analysis we can highlight the sustained part of each of the harmonics. This helps the audio expert who usually selects by hand the ADSR segments during the sound design process.

The sound synthesis analysis results are shown from Fig. 7 through Fig. 10. Fig. 7 illustrates the Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) averaged along the time axis versus the sound synthesis parameter (decreasing number of clusters). It shows that as the number of clusters is reduced (less resources used at the sound synthesis stage), the error increases. Fig. 8 shows a surface plot of the difference waveform error. Both give very poor real indication about the sound synthesis quality.

In Fig. 9, an ODG curve (averaged over all) showing minimal / mean / maximal values versus the number of clusters is shown. It indicates the performance of the synthesis technique and sound quality along the sound design process. As the number of clusters used in the synthesis decrease the sound synthesis quality decreases too. This curve could be broadly divided into three parts. The first part in which degradation increase slowly, then the degradation falls quickly and stabilizes towards the end.

The most interesting results are shown in Fig. 10. It presents an "ODG surface". The y-axis corresponds to the variation of a sound synthesis parameter (i.e. the number of clusters used in the synthesis) and x-axis is the time axis (frames). We have added a "perceptible threshold plan" (i.e. ODG equal to -1) that indicates the level at which PEAQ can detect that the degradation starts to be perceptible.

4.2. Flute Sound

The second example is a flute sound. The sound analysis engine estimates the fundamental to be 527.473 Hz. This sound has very clean timbre. The sound analysis resulted in 46 harmonics. Only one harmonic seems dominant and gives this sound a very sine wave like characteristic. The sound waveform is shown in Fig. 11, the sound spectrum in Fig. 12 and harmonics analysis in Fig. 13.

Results of the analysis of sound synthesis quality are show in Fig. 14 and Fig. 15. Reducing the number of clusters reduces the perceived sound quality. The average ODG (avgODG) curve presents a flat characteristic until it drops gently with a linearly characteristic. This indicates that the algorithm start reducing clusters with no perceptual degradation. The variation of between minimal / maximal values of ODG in Fig.14 is much more smaller than in previous example. Fig. 15 shows the ODG surface.

4.3. Principal Stop sound

The last example is a Principal Stop sound. The fundamental of this sound is 217.195.84 Hz. The sound analysis resulted in 111 harmonics.

The full analysis includes the sound waveform shown in Fig. 16, the sound spectrum in Fig. 17 and harmonics analysis in Fig. 18. Few harmonics have very stable behavior (compared to Choir sound example) giving this sound a defined characteristic.

Results of the analysis of sound synthesis quality are show in Fig. 19 and Fig. 20. Reducing the number of clusters reduces the perceived sound quality, in a more progressive way in this case.

Overall results show that only impairments between "imperceptible" and "perceptible, but not annoying" were considered acceptable by the audio experts. Visual inspection of the results in 3D (perceived audio quality versus synthesis parameter versus time) allows us to correlate the perceived audio quality with the sound attack, sustain, and release timing segmentation.

They also gives indications about the progression in the sound design process, and can be used to reveal imperfections of a sound synthesis technique and clustering rules in the case of our intelligent system. This is greatly helping to refine the rule-based fuzzy expert system and fine-tune the fuzzy sets and rules of the fuzzy expert system used for as the Knowledge-based Audio Feature Processing Engine.

More audio examples and detailed results will be made available, before the convention, at the project home page [9]. .

5. CONCLUSIONS

In this paper, we have presented the objective prediction of sound synthesis quality looking at the final part of our intelligent audio system for perceptual-based optimization of sound synthesis. Audio experts' listening tests are now supported by a perceived sound quality assessment based on the ITU-R BS1387 Perceptual Evaluation of Audio Quality (PEAQ) algorithm.

Results from pipe organ sound database shows that plots of perceived sound quality versus number of clusters gives good indications about the progression in the modeling process, and can be used to reveal imperfections of a sound synthesis technique. It is used to improve clustering rules in the case of our intelligent system. This has proven to greatly help to refine the rule-based fuzzy expert system and fine-tune the fuzzy sets and rules of the fuzzy expert system engine.

Results also indicate that 6 to 12 clusters are often required for modeling the majority of pipe organ sounds, depending on the pitch of the sound and the type of pipe. This number can decrease to low as one or two for sounds with high pitch.

The system helps the audio experts to quantify the perceived sound quality of the synthesized sounds. It serves as a support tool, and helps to reduce time-consuming listening tests. However, more work is still needed to fully exploit the potential of objective measures of perceived audio quality in sound synthesis.

Future work should take into consideration the specificity of the PEAQ algorithm's cognitive model (audio coding artifacts as describes by Erne in [10]) and quality assessment of audio experts considered as having “Golden Ears” [11][12] when developing a new metric for sound quality index an expert level.

In future, we plan to investigate the use of knowledge gained using PEAQ as a basis for the development of a new index to accurately and objectively predict sound quality to benchmark sound synthesis techniques such as additive synthesis, wavetable synthesis, frequency modulation synthesis and digital waveguide modeling, with application to musical instruments such as piano [13] and church bells [14].

6. ACKNOWLEDGMENTS

This research is supported by the Engineering and Physical Science Research Council (EPSRC) under the grant (GR/S00859). We are grateful to Tony Koorlander and Graham Blyth for their help during the course of this project, and fruitful discussions. 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).

7. REFERENCES

[1] E. Zwicker and H. Fastl, Psychoacoustics, Facts and models. Springer Verlag, 1999.

[2] B. Hamadicharef and E. C. Ifeachor, “An Intelligent System Approach To Sound Synthesis Parameter Optimisation,” presented at the AES111th convention, New York, USA, 2001 November 30 – December 3.

[3] T. Koorlander, Personal email communications, 2002.

[4] ITU-R BS.1116,. "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems". 1994-1997

[5] 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,” J. Audio Eng. Soc., vol. 48 pp. 3-29, 2000.

[6] ITU-R BS.1387, "Method For Objective Measurement of Perceived Audio Quality," 1998.

[7] Opticom, 2001. “OPERA: Voice/Audio Quality Analyser”. Brochure and User Manual Version 3.5.

[8] A. Lerch, Personal email communications, 2002.

[9] http://www.tech.plymouth.ac.uk/spmc/S00859/

[10] M. Erne, “Perceptual Audio coders: What to Listen For,” presented at the AES111th convention, New York, USA, 2001 November 30 – December 3.

[11] S. Shlien, and G. Soulodre, “Measuring the Characteristics of “Expert” Listeners,” presented at the 101st Audio Eng. Soc., Los Angeles, USA, 1996, November 8-11.

[12] S. Shlien, “Auditory Models for Gifted Listeners,” in J. Audio Eng. Soc., vol. 48, pp. 1032-1044, November 2000.

[13] J. Laroche, and J. L. Meiller, “Multichannel Excitation/Filter modeling of percussive sounds with application to the Piano,” in IEEE Transactions on Speech and Audio Processing, vol. 2, pp. 329-344, IEEE, April 1994.

[14] M. Karjalainen, V. Välimäki, and P. A. A. Esquef, “Efficient Modeling and Synthesis of Bell-like Sounds,” in Proc. of the 2002 Conference on Digital Audio Effects, pp. 181-186, DAFX, September 2002.

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 with meaning

Model Output Variables	Purpose
WinModDiff1	Changes in modulation
AvgModDiff1	(related to roughness)
AvgModDiff2	
RmsNoiseLoud	Loudness of the distortion
BandWidthRef	Linear distortions
BandWidthTest	(frequency response, etc.)
RelDisFrame	Frequency of audible distortions
Total NMR	Signal-to-mask ratio
MFPD	Detection probability
ADB	Detection probability
EHS	Harmonic structure of the error

Table 2: Model Output Variables (MOVs)

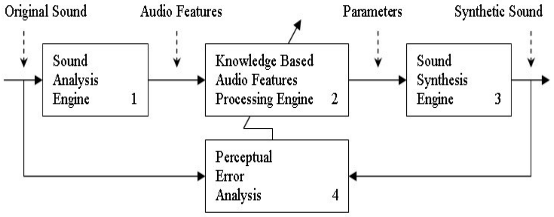
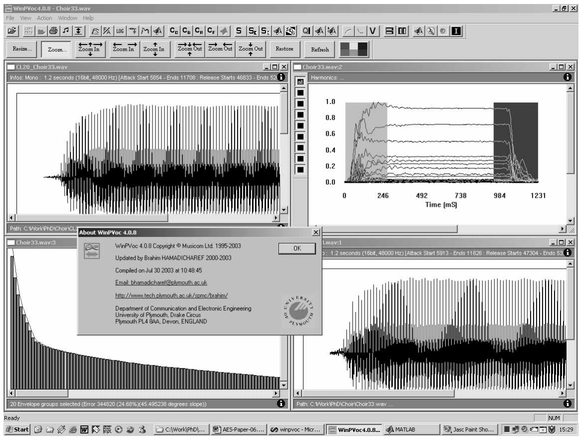


Fig. 1: Perceptual-based optimization of sound synthesis


About WinPVoc 4.0.8

WinPVoc 4.0.8 Copyright © Musicom Ltd. 1995-2003

Updated by Brahim HAMADICHAREF 2000-2003

Compiled on Jul 30 2003 at 10:48:45

Email: bhamadicharef@plymouth.ac.uk

http://www.tech.plymouth.ac.uk/spmc/brahim/

Department of Communication and Electronic Engineering
University of Plymouth, Drake Circus
Plymouth PL4 8AA, Devon, ENGLAND

OK

20 Envelope groups selected (Error 344820 (24.68%)(45.495238 degrees slope))

Fig. 2: Screenshot of the software tool

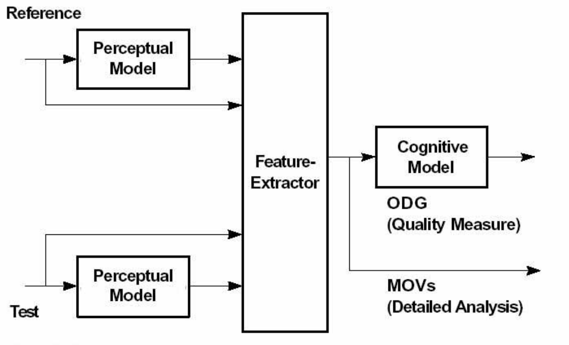


Fig. 3: Generic perceptual measurement algorithm

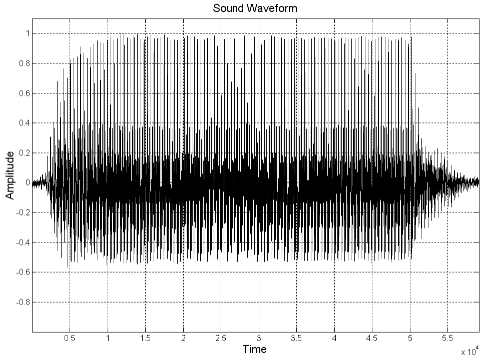


Fig. 4: Choir waveform sound analysis

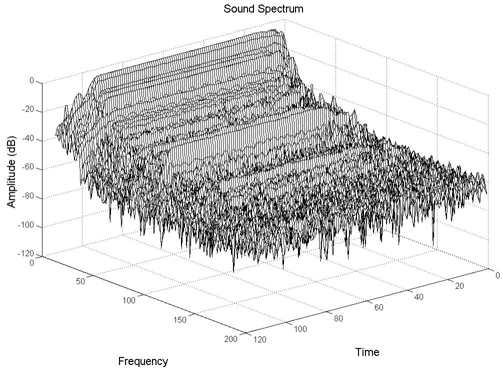


Fig. 5: Choir sound spectrum analysis

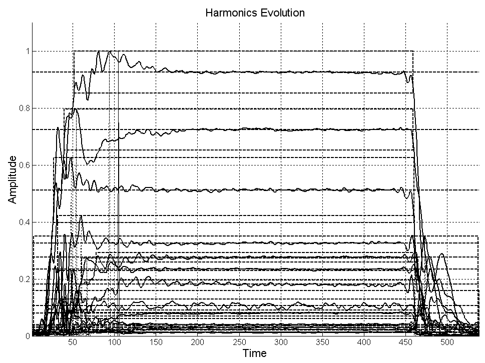


Fig. 6: Choir sound harmonic analysis

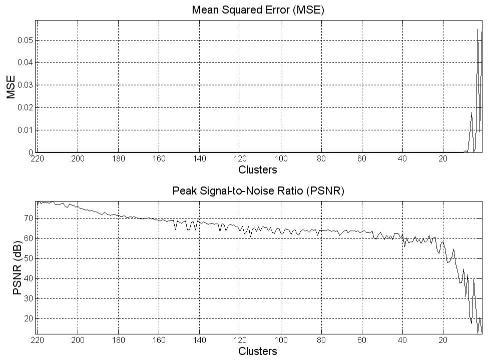


Fig. 7: Choir sound synthesis MSE / PSNR error

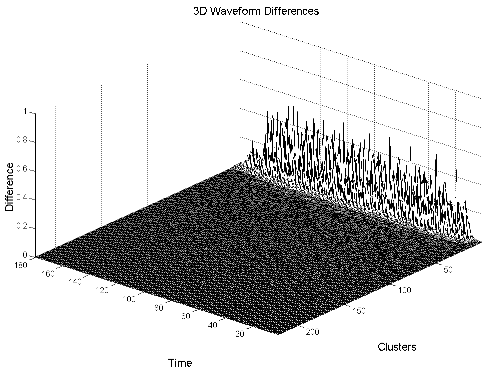


Fig. 8: Choir waveform difference error

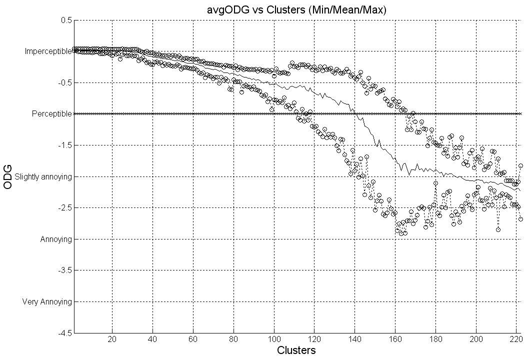


Fig. 9: Choir ODG versus clusters

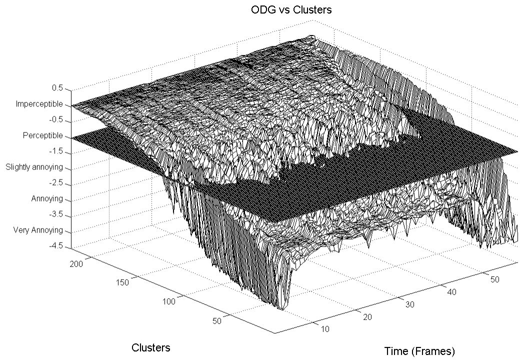


Fig. 10: Choir ODG surface

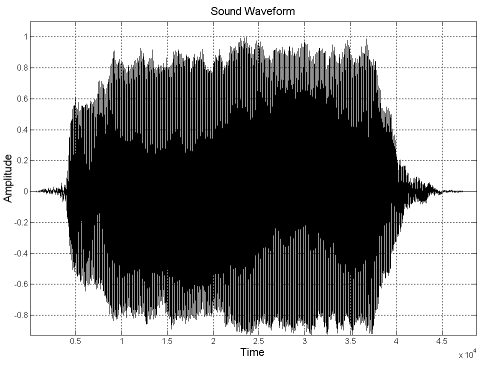


Fig. 11: Flute sound waveform analysis

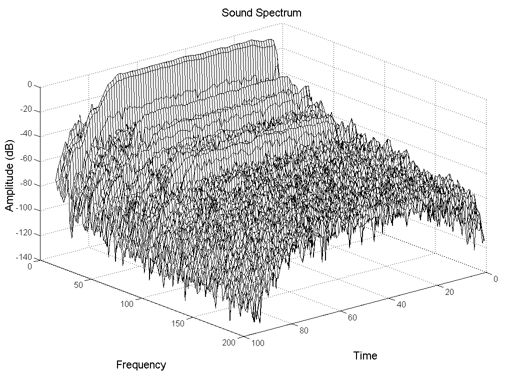


Fig. 12: Flute sound spectrum analysis

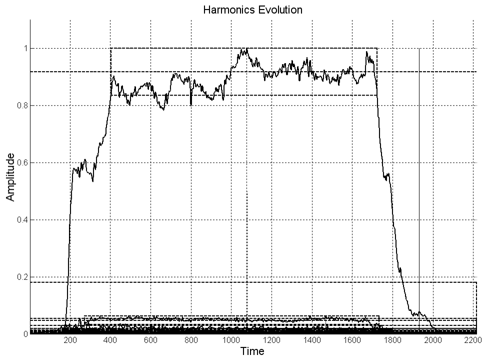


Fig. 13: Flute sound harmonic analysis

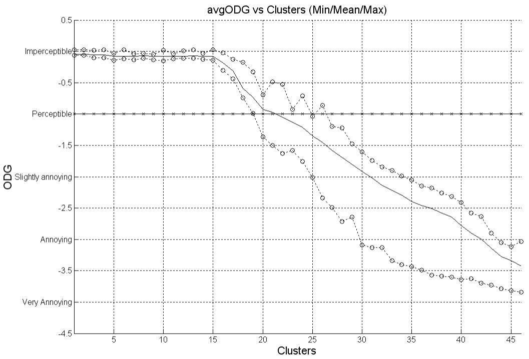


Fig. 14: Flute ODG curve

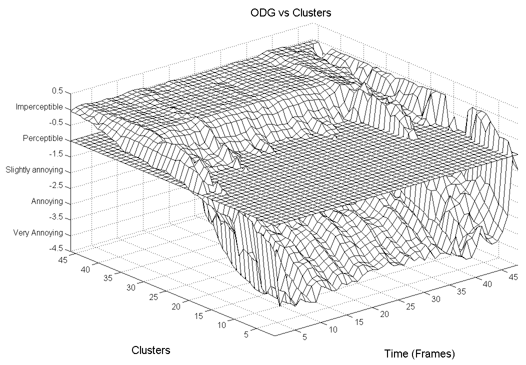


Fig. 15: Flute ODG surface

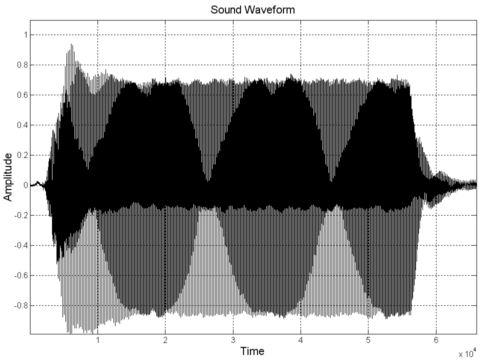


Fig. 16: Principal Stop sound waveform analysis

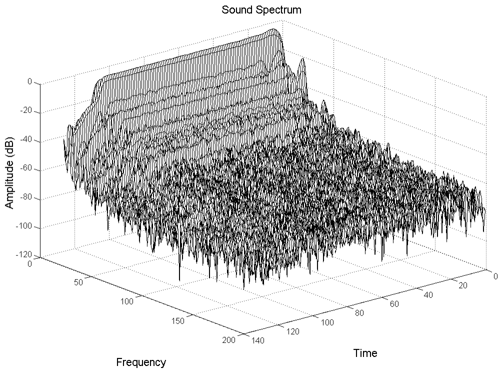


Fig. 17: Principal Stop sound spectrum analysis

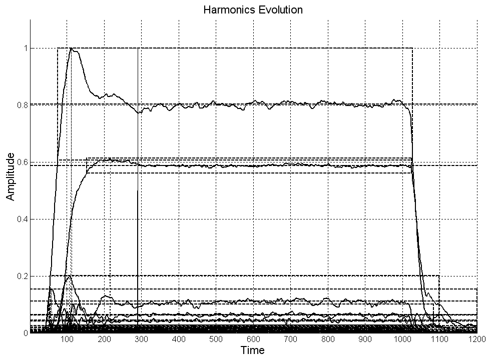


Fig. 18: Principal Stop sound harmonic analysis

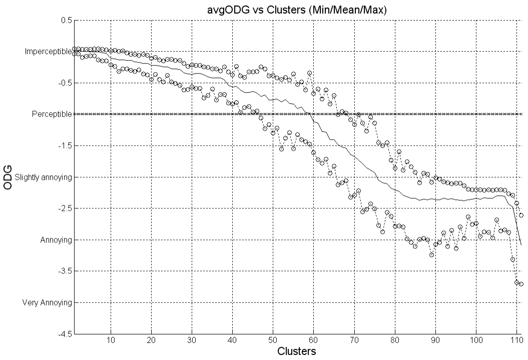


Fig. 19: Principal Stop ODG curve

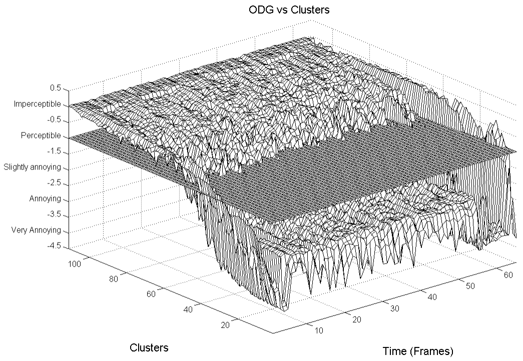


Fig. 20: Principal Stop ODG surface