Contents lists available at SciVerse ScienceDirect



Expert Systems with Applications



journal homepage: www.elsevier.com/locate/eswa

Intelligent and perceptual-based approach to musical instruments sound design *

Brahim Hamadicharef*, Emmanuel C. Ifeachor

Signal Processing and Multimedia Communication (SPMC) Research Group, School of Computing and Mathematics, Faculty of Science and Technology, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

ARTICLE INFO

Keywords: Sound design Audio expertise Modeling musical instruments

ABSTRACT

A novel, generic framework for musical instrument sound design is proposed. The framework uses an intelligent and perceptual based approach to address the two main problems in sound design – optimization of synthesis parameters and assessment of sound quality. A fuzzy model is used to capture and exploit knowledge of sound design from audio experts. A robust methodology, based on the ITU Perceptual Evaluation of Audio Quality (PEAQ) algorithm, is used for objective prediction of sound synthesis quality and sound quality assessment. The new framework makes it possible to automate the optimization of synthesis parameters. It also allows the designer to evaluate objectively, the perceptual impact of individual parameters on the final sound quality which is useful for benchmarking sound synthesis methods. The framework is generic and can be used for sound design for a wide range of musical instruments. We illustrate the use of the framework in pipe organ sound design. Results from this show that the new approach provides an important and useful alternative to existing methods.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Design and synthesis of musical instrument sounds are key aspects of the growing field of computer music technology (Lee & Horner, 1999). Sound design involved the process of sound analysis, processing, synthesis and quality assessment. It makes possible the creation of new music well beyond the physical boundaries of traditional acoustic musical instruments. It is an important research field which informs the design and use of electronic musical instruments (Smith, 1996) and aids the preservation of historical instruments (Rioux, 2000) and cultural heritage.

However, despite advances in signal processing, sound design for complex musical instruments, such as pipe organs, require a deep understanding of the instrument (e.g. how the resonance of pipes contribute to the dynamic pitch variation and character of the overall sound and what sound features to use during synthesis to retain these) and significant expertise to produce synthetic sounds of sufficient quality to meet the needs and desires of musicians and musical instrument builders. Some of the best results are still found empirically and the process is time-consuming because of the enormous complexity and size of the data structures (Sandell & Martens, 1995).

* Corresponding author. Tel.: +65 98610618.

There are two main limitations in sound design and synthesis. First, sound analysis requires skills and understanding only gained over many years. In practice, the data needs to be represented in a form that would allow easy manipulation using parameters that an instrument builder or musician would easily relate to and understand. Semi-automatic analysis software tools exist to assist sound design but these also require skilled user intervention in order to achieve acceptable results. The process is time consuming and error prone, severely restricting productivity. The modeling of a complete instrument, such as a pipe organ, is difficult because of its complexity (e.g. the different arrangements of stops, each with different sonic characteristics, the effects of the shape and construction of the pipe Hopkins & Rimbault, 2000). The task is further complicated by the fact that in a piece of music, several notes may be played simultaneously and the combination of sounds is important in order to re-create the musical experience as close to the original as possible.

A number of methods exist for sound design and musical instrument modeling. Group Additive Synthesis (GAS) (Kleczkowski, 1989; Oates & Eaglestone, 1997) may be used to optimize sound synthesis using a data-reduction technique by identifying the redundancy within the harmonics envelopes and clustering these into groups resulting in one composite amplitude and frequency envelope per group. The performance of GAS methods are affected by factors such as the choice of the method used to calculate the similarity and distances between harmonics. The harmonics of very complex sounds cannot be easily be clustered, resulting in poor synthetic sound quality. Wavetable matching using Genetic Algorithm (GA) (Goldberg, 1989) provides an important alternative

 $^{^{\}star}$ This work was supported by the Engineering and Physical Science Research Council (EPSRC) under the Grant GR/S00859/01.

E-mail addresses: bhamadicharef@hotmail.com (B. Hamadicharef), e.ifeachor@ plymouth.ac.uk (E.C. lfeachor).

^{0957-4174/\$ -} see front matter \odot 2011 Elsevier Ltd. All rights reserved. doi:10.1016/j.eswa.2011.12.058

approach (Cheung & Horner, 1996). Typically, time–frequency analysis is first performed using, e.g. a Phase-Vocoder (Dolson, 1986), to extract time-varying evolution of the sound harmonics. Then, GA is used to search and optimize the grouping of the harmonics into wavetables, which are recombined to create the tones. This approach has been successfully applied to a number of musical instruments (Wun & Horner, 2006).

The second limitation is related to sound quality assessment. The synthetic sound needs to be auditioned, using listening tests, and synthesis parameters manipulated and refined to achieve the best results. Currently, the preferred method for sound quality assessment is listening tests. These are typically carried out by listeners with good background in sound synthesis and music (So & Horner, 2002). Listening tests are expensive, time consuming and require specialized sound facilities (ITU-R Recommendation BS.1116, 1997; ITU-R Recommendation BS.562-3, 1997). Furthermore, in an automated sound design procedure, they are impractical. In more recent studies, objective metrics such as the mean relative error (MRE) is used. It is defined as:

$$MRE = \frac{1}{N_{FRM}} \sum_{n=1}^{N_{FRM}} \sqrt{\frac{\sum_{k=1}^{N_{HAR}} \left[A_k(n) - A'_k(n)\right]^2}{\sum_{k=1}^{N_{HAR}} A_k(n)^2}}$$
(1)

where A_k and A'_k are the amplitude of the *k*th harmonics of the original and synthetic sound respectively, N_{FRM} is the number of frames and N_{HAR} the number of harmonics.

This metric is used to measure the difference between the original and synthetic sounds (Horner, Beauchamp, & So, 2006) and as a measure of goodness for the selection process in the GA-based optimization procedure (Wun & Horner, 2005). The use of objective metrics such as the MRE is important, but perceptual based methods provide a better objective measure of the perceived sound quality. In sound design, the quality of the audio is affected by the nonlinear operations involved in sound analysis, feature processing and sound synthesis. The effect of such operations are complex and cannot be adequately measured using conventional techniques as they fail to reveal quality as perceived by end users. Perceptual methods are based on models of the human auditory system and thus provide a direct and closer link to audio quality as perceived by the user.

The main contributions of this paper are:

- 1. A novel, generic framework that provides an effective solution to the two main limitations in sound design. The framework is based on an intelligent and perceptual-based approach, combining a fuzzy model with an objective sound quality assessment methodology.
- 2. A fuzzy model for capturing the skills and knowledge from pipe organ experts for sound design and sound synthesis optimization.

- 3. A sound quality assessment methodology based on the Perceptual Evaluation of Audio Quality (PEAQ) algorithm, the ITU-R Recommendation BS.1387 (ITU-R Recommendation BS.1387, 1998) which can be used to automate the sound design process, removing the need for listening tests.
- 4. An illustration of the use of the framework, fuzzy model and audio quality assessment method in the design of sound quality for pipe organ.

The remainder of the paper is organized as follows. In Section 2, we describe the generic framework for sound design, including the fuzzy model of pipe organ experts and methodology for objective prediction of sound synthesis quality. The use of the framework in pipe organ sounds design, including practical results, are illustrated in Section 3. Finally, we conclude the paper in Section 4.

2. New framework

2.1. Overview

The block diagram in Fig. 1 depicts the new framework for sound design. It consists of four main stages: sound analysis, sound features processing, sound synthesis and sound quality assessment. Starting with an input waveform, with stereo converted into monophonic sounds, the sound analysis stage performs a time–frequency analysis to produce temporal and spectral sound features. These are processed, by a fuzzy model of audio expertise, at the sound features processing stage to generate the best possible synthesis parameters (in terms of sound quality). The parameters are then used to re-create the original sound. Finally, the sound quality assessment uses the original and synthetic sounds to calculate an index of quality to give an indication of how good the resulting sound is compared to the original.

The framework is generic and can use any type of sound analysis and synthesis methods and for sound design of most musical instruments. For analysis, typically Fourier-based methods (e.g. the Phase-Vocoder Dolson, 1986) are used. The sound synthesis method used depends on the type of instrument. Typically, for piano we use digital waveguide (Smith, 1992) and excitation/filter modeling (Laroche & Meillier, 1994), and for pipe organ we use additive synthesis (Comerford, 1993) and wavetables (Horner, Beauchamp, & Haken, 1993).

The framework has three different modes: (1) a modeling mode in which it is used to re-create synthetic sounds based on the fuzzy model of audio expertise, (2) an optimization mode in which sound synthesis parameters are tuned and the final quality monitored for best result, and (3) an evaluation mode in which individual synthesis parameters are adjusted to assess their perceptual impact on the final perceived sound quality. All three modes support



Fig. 1. Block diagram of the intelligent and perceptual-based sound design system.

investigations into important and most time-consuming aspects of sound design and musical instrument modeling before the choice of implementation and hardware resources is made.

2.2. Fuzzy model of audio expertise

A fundamental problem in the development of a model of audio expertise to process the audio features extracted by the sound analysis engine is how to handle imprecision and uncertainty in sound design. Fuzzy Logic is used as it offers a comprehensive and flexible framework for handling the imprecision and uncertainty that characterize audio knowledge, data and decision-making (Zadeh, 1983). In particular, it provides a framework for describing, manipulating, conveying information and drawing conclusions using linguistic terms as in the real world. Thus, the linguistic terms (or variables) and rules which the music technologists use and understand can be used directly. This makes the model accessible in a natural form which is an important factor in the successful development of the model of audio expertise (Garibaldi & Ifeachor, 1999).

The two key concepts in fuzzy logic are linguistic variables and fuzzy sets. Linguistic variables are subjective, context-dependent variables whose values are words. For example, if the phrase "attack time" is regarded as a linguistic variable, the values could be *very fast, fast, slow* and *very slow* and can be denoted as – Attack-Time (very fast, fast, medium, slow, very slow) (see Fig. 6). Fuzzy sets representing linguistic variables can use different types of



Fig. 2. Different shapes of fuzzy sets membership functions.

membership functions (e.g. Sigmoid, trapezoidal, Bell shape and triangular) as shown in Fig. 2. Fuzzy rules are typically in the form of IF-THEN:

$$IF [x_1] is [S_1] AND \dots, [x_n] is [S_n]$$

$$THEN [y_1] is [T_1] AND \dots, [y_m] is [T_m]$$
(2)

where $[y_i]$ is the output fuzzy variables (consequents) whose values are inferred, $[x_i]$ is an input fuzzy variable (antecedent), $[S_n]$ and $[T_m]$ are input and output fuzzy sets. Examples of such rules for pipe organ sound design are given in Section 3.1. To apply the model involves three basic operations, as shown in Figs. 3 and 4. First, the input variables, *i.e.* sound features, must be converted into fuzzy variables using the membership functions. Then, the fuzzy inference engine processes the variables by applying the rules. Essentially, for each rule in turn, the membership grade for each fuzzy set is evaluated by determining the value of membership function at the given value of the corresponding input variable. Membership grades are then combined by a suitable fuzzy operator (e.g. min and max) to give the extent the rule will be activated and used to truncate the associated rule consequent fuzzy set. The consequences of all rules that activated are combined to give an overall fuzzy consequent. The fuzzy consequence is then defuzzified to produce a crisp output for the outside world (in our case parameters to control sound synthesis). The development of the model follows the stages of (1) definition of the problem with experts during knowledge elicitation sessions, (2) definition of the fuzzy sets and fuzzy variables used in the fuzzification and defuzzification processes, and (3) development of the rules and inference. The model is then evaluated and refined.

2.3. Objective prediction of sound synthesis quality

An important aspect of sound synthesis is the assessment of the perceptual impact of individual parameters on the final sound quality. We developed an objective methodology for sound synthesis quality prediction (Hamadicharef & Ifeachor, 2003) which exploits the Perceptual Evaluation of Audio Quality (PEAQ) algorithm (ITU-R Recommendation BS.1387, 1998; ITU-R Recommendation BS.1115, 1993) in the sound design process.

As shown in Fig. 5, PEAQ consists of three main stages (ITU-R Recommendation BS.1387, 1998): a psychoacoustic model (a combination of a peripheral ear model and pre-processing of excitation patterns stage), a feature extractor and cognitive model. PEAQ takes an audio reference and test files as inputs and produces quality measures. Internally, the psychoacoustic models are applied to both reference and test signals to extract excitation patterns such as loudness, modulation and error. These patterns are then pre-processed to calculate the Model Output Variables (MOVs). MOVs



Fig. 3. Typical fuzzy rule-based model.



Distortion Index (DI)

Objective Difference Grade (ODG)

Fig. 5. Perceptual Evaluation of Audio Quality (PEAQ) algorithm (ITU-R Recommendation BS.1387, 1998; Keyhl et al., 1999).

 Table 1

 Objective Difference Grades (ODG) and their meaning.

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

include loudness of distortion, changes in modulation, linear distortions, harmonic structure of error, noise-to-mask ratio (NMR). The MOVs are mapped into quality measures: the Distortion Index (DI) and the Objective Difference Grade (ODG). This mapping uses an Multi-Layer Perceptron (MLP) trained on the ITU databases (Thiede et al., 2000; ITU-R Recommendation BS.1115, 1993). As shown in Table 1, ODG ranges from 0.0, indicating a measured audio quality with imperceptible impairment, down to -4.0, for very annoying impairment. In this study, we are interested in the ODG = -1.0, which corresponds to perceptible but not annoying impairments. All the assessments performed were carried out with an Opera *Voice/Audio Quality Analyzer* (Opticom GmbH, Germany) (Keyhl, Schmidmer, & Wachter, 1999).

A basic version and an advanced version of PEAQ exist. The psychoacoustic model of the basic version uses a Discrete Fourier Transform (DFT) to analyze the audio signal and generates 11 MOVs, while the advanced version uses both a DFT and a filter bank and produces only 5 MOVs. The MLP of the basic version has a [11-3-1] topology while for the advanced version it has a [5-1] topology (no hidden layer).

3. Application to pipe organs

In this section, we illustrate the use of the framework in pipe organ sound design (Hamadicharef & Ifeachor, 1999).

3.1. Modelling pipe organs sound design expertise

3.1.1. Knowledge elicitation

Expert knowledge was collected and formulated for the development of the fuzzy model. Over the course of the project, knowledge was elicited continually and reviewed with a pipe organ expert.¹ Different elicitation techniques were used to obtain sufficient knowledge necessary for the design, development of the fuzzy model. These include formal and informal interviews, correspondence by emails, as well as from the literature on musical instruments (Fletcher & Rossing, 1998), pipe organs (Hopkins & Rimbault, 2000) and psychoacoustics (Zwicker & Fastl, 1999).

3.1.2. Development of a fuzzy model

The fuzzy rule-based model was developed following the description in Section 2.2. The fuzzy sets were developed for each sound feature. An example of such fuzzy sets is shown in Fig. 6 with the Attack Time as a variable. It consists of five fuzzy subsets: *VeryFast, Fast, Medium, Slow* and *VerySlow*. If a very long attack time is considered to be 100 ms, then the points at which the fuzzy sets cross the universe of discourse (horizontal axis) correspond to the following time intervals – 0 to 3 ms for a *VeryFast* attack, 3–12 ms for a *Fast* attack. An attack time was considered to be *Normal* if it is between 12 and 30 ms, whereas a *Slow* attack would be between 30 and 80 ms, and finally a *VerySlow* attack time would be 80 ms and above.



Fig. 6. Example of fuzzy set: attack time.

From the knowledge acquired, a set of rules for sound design/ synthesis process was produced. An example of a rule set for organ pipe sounds (in this case a rule set for the overall shape of the envelopes of harmonics in terms of amplitudes) is:

"IF the attack of the harmonics are short THEN the overall sound is PERCUSSIVE";

"IF the variations of the harmonics in sustain are medium THEN the sound is EVEN";

"IF the variation of the harmonics in sustain are very large THEN the sound is UNTIDY".

It was found important to represent rules in the form of IF-THEN statements and with linguistic terms that musicians/musical instrument builders can relate to. Finally, because experts understand theses rules, this also facilitated feedback and enabled them to actively participate in the development of the system.

Examples of terms used to describe pipe organ sounds are shown in Table 2. To illustrate, a *flutev* sound is considered as a soft sound with predominant 1st harmonic and a limited range of about 10 harmonics decreasing rapidly in amplitude beyond the 5th. A *bright* sound is a *clear* sound with harmonics that extend into the upper frequencies of about 12–15 kHz. A slow sound has a low frequency note with harmonics that take longer than 50 ms to start. Typically, nasal sounds have louder harmonics in the 3rd to 9th range than at the 1st harmonic. A breathy sound has air noise that is about the same amplitude as the harmonics. It can also be a sound which has a very unstable upper order harmonic structure and unstable in terms of pitch or amplitude. A sound can be defined as harsh when typically the harmonic amplitudes stay the same right up to say the 50th harmonic, with very little sustain variation. The 1st to 5th harmonics may often be lower in amplitude than the others. If the variations of the harmonics in the sustain part are large then the sound is said to be untidy. When the variations of the harmonics in the sustain part are medium then the sound is said to be even. A sound is often said to be percussive if the attack part of its dominant harmonics are short.

The synthesis parameters are optimized in relation to timing, amplitude, frequency, amplitude and frequency modulation, and noise, all of which are important in sound design and modeling musical instruments. Four distinct fuzzy outputs representing key aspects of the sound synthesis parameters were defined. They are the *cluster*, the *attack*, the *sustain*, and the *noise*. The *cluster* fuzzy output provides an indications about which cluster the harmonic should belong to. In a similar manner to experts designing sounds, the fuzzy model is used to assess the perceptual contribution of each harmonic to the final sound quality and to group

¹ Mr. Tony Koorlander (Musicom Ltd., bideford, UK).

Table 2	
Linguistic descriptors used by pipe organ experts.	

Description	Linguistic terms
General impressions	Old, noisy, pleasant, relaxed, simple, stable, strong, tensed, thin, undefined, unfocused, unpleasant, unstable, warm, weak
Transient part	Aggressive, strong, gentle, long, weak, sounds like chiff, short, sounds like cough, fast, slow, sounds like hiss, soft, connected, disconnected, integrated, related
Steady state	Airy, breathy, bright, clean, clear, cold, dirty, dull, floppy, flowy, fluffy, flutey, free, full, harsh, horn-like, leaky, loose, nasal, reedy, oppressive, rough, round, sandy, sharp, singing, splitting, stringy, thin

harmonics into clusters, isolating them based on their characteristics. The *attack* fuzzy variables control the raising of the amplitude envelopes and small variations of the amplitude modulations using parameters in form of an amount and variation rate. These parameters are considered important by the audio experts in order to obtain realistic and high quality pipe organ sounds. The *sustain* fuzzy variables control the envelopes and modulations both amplitude and frequency in the sustain part of the sound. Modulations in amplitude and frequency represent key issues in sound design and are dealt with by adding small amounts of amplitude and frequency modulations to harmonics envelopes, adding "life" to the final sound (Ando & Yamaguchi, 1993). In general, the attack section is being dealt with after the sustain part is considered satisfactory. A *noise* fuzzy variable controls specific noise sources added



Fig. 7. Time-varying envelopes of each harmonic for (a) the Oboe 8', (c) the Metal trombone 16' and (e) the Dulciana 16'. Harmonic distribution for (b) the Oboe 8', (d) the Metal trombone 16' and (f) the Dulciana 16'.

the sound to improve the overall realism and emulate, for example, typical pipe organ's characteristic such as the *chiff*.

Following discussions with audio experts, extensions to the above model are being developed. These involve smaller decision-making models related to each family of pipe, with some refinements on the basis of individual stop. In such context, a global fuzzy model of pipe organ can be thought of as the combination of smaller models. Each pipe, due to its physical construction (Hopkins & Rimbault, 2000) has its own characteristic and thus specific fuzzy models would be more appropriate. Rules dealing with specific rank/stops/pipes issues can be easily incorporated into a global model for a pipe organ instrument. Ultimately, they will be part of the sound design tools used by electronic pipe organ manufacturers.

3.2. Intelligent sound design

We will describe the step-by-step procedure. As described in the previous sections, the system takes a sound recording as input (with stereo recordings first converted to mono) and generates synthetic sounds following the four stages detailed in Fig. 1. First the sound is analyzed using Phase Vocoder to extract sound features (this may be proceeded by some pre-processing steps such as normalization, background noise reduction, etc.). The sound features consist of attack, decay and release times, the amplitude envelope with its modulations, the distribution of the harmonics, the frequency envelope and its modulations, the pitch, some modified Tristimulus (Kostek & Czyzewski, 2001), defined as:

$$T_1 = \frac{A_1}{\sum_{k=1}^{N} A_k^2}$$
(3)

$$T_2 = \frac{\sum_{k=2}^4 A_k^2}{\sum_{k=1}^N A_k^2} \tag{4}$$

$$T_{3} = \frac{\sum_{k=5}^{N} A_{k}^{2}}{\sum_{k=1}^{N} A_{k}^{2}}$$
(5)



Fig. 8. Average ODG for (a) the Oboe 8', (c) the Metal trombone 16' and (d) the Dulciana 16'. Surface ODG for (b) the Oboe 8', (d) the Metal trombone 16' and (f) the Dulciana 16'.

contents of even (T_{even}) and odd (T_{odd}) harmonics in the sound spectrum, defined as:

$$T_{even} = \frac{\sqrt{\sum_{k=1}^{M} A_{2k}^2}}{\sqrt{\sum_{k=1}^{N} A_k^2}}$$
(6)

$$T_{odd} = \frac{\sqrt{\sum_{k=1}^{M} A_{2k-1}^2}}{\sqrt{\sum_{k=1}^{N} A_k^2}}$$
(7)

where A_k is the amplitude of the *k*th harmonic, *N* the number of harmonics and $M = \lfloor (N/2) \rfloor$. The brightness (Br) of a sound, or spectral centroid, is defined as:

$$Br = \sum_{k=1}^{N} kA_k \bigg/ \sum_{k=1}^{N} A_k$$
(8)

The attack section was given a particular attention as it is well known in sound synthesis that it contributes to the most of the naturalism and realism of the sound (Ando & Yamaguchi, 1993). Sound features are processed by the fuzzy model which generates parameters for the sound synthesis stage, to recreate the synthetic version of the original sound.

We will illustrate the use of the framework in the design of three pipe organ sounds. Each is taken through the process of sound analysis, feature processing/optimization, sound synthesis and quality evaluation and prediction. These three sounds were selected by the experts as they were considered to be challenging examples. They were taken from a large sound database of recordings made in England (e.g. Hexham Abbey, Devon, England) and United States (e.g. First Presbyterian Church of Kilgore, Texas, USA).

The first sound is from an Oboe 8' (Key G2, with a fundamental of about 197 Hz), the second was recorded from a Metal trombone 16' pipe with a fundamental of about 65 Hz (key C1), and the third sound is from a Dulciana 16' with a fundamental of about 131 Hz (key C2). Time-varying evolution of each harmonic and overall harmonic distribution (only the first 16 harmonics) are shown in Fig. 7(a) and (b) for the Oboe sound analysis, in Fig. 7(c) and (d) for the Metal trombone 16', and in Fig. 7(e) and (f) for the Dulciana 16'.

The sound analysis reveals some very slow attack harmonics for the Oboe sound, while the attack of the Metal trombone shows clear instability. The attack of the Dulciana sound has obvious significant overshoot. This can be seen from the sustain level (automatically estimated by the system) shown on the bar graph and as dotted line on the time-varying harmonic envelopes. The harmonic distribution provides a very good account of the harmonic structure and dominance. For example it can be seen that the Dulciana sound's main harmonic is the 2nd, while for the two others the 6th is the highest. Harmonics of the Oboe from 8th upwards are very weak, while for the Metal trombone many still have a lot of energy.

We also present results from the evaluation of the perceptual impact of sound design (using clustering) parameters which were found useful. We implemented and evaluated the Agglomerative Hierarchical Clustering (AHC) method (Oates & Eaglestone, 1997) with various distance used for the agglomeration stage of the clustering process. We investigated Euclidean, maximum, Manhattan or City Block metric metrics methods, as well as Universal Quality Index (UQI) (Wang & Bovik, 2002).

Results showing the impact of clustering reduction during sound design are important for expert to evaluate sound resource requirements and optimization. The motivation behind such evaluation was to find out the perceptual threshold at which parameter values can be adjusted until ODG values are below the -1.0

(corresponding to the imperceptible audio impairment). Such measure would be more objective than the vague term "indistinguishable" used in other studies (Horner et al., 2006) with listening tests.

Fig. 8(a) (for the Oboe 8'), Fig. 8(c) (for the Metal trombone 16') and Fig. 8(e) (for the Dulciana 16') show results from an evaluation of the impact of clustering during sound design. The sound quality impairments clearly decrease over time (minimal, average and maximal of ODG value are shown). We were also interesting in a 3-dimensional view to relate such issue to time. The most important aspect was to look at the intersection between the *quality surface* with the *imperceptible plan* (ODG = -1.0). Results are shown for the Oboe, the Metal trombone and the Dulciana in Fig. 8(b), (d) and (f), respectively.

The feedback from users of the new approach was positive and proved that it is useful as support tool for sound designers. Our approach is promising but requires more work focused on specific organ pipes. Furthermore, experiments have shown that perceptual threshold from PEAQ ODG values do not strictly correspond to the same level from pipe organ expert's quality assessment. Further work is currently investigating this aspect with the study of PEAQ's cognitive model using synthetic pipe organ sound database.

4. Conclusion

In this paper, we have presented an intelligent and perceptualbased approach to musical instrument sound design. We have proposed a novel sound design framework to address the two major limitations in sound design: the synthesis parameter optimization and assessment of sound quality. With the aid of audio experts, we have developed a fuzzy model to capture and exploit the expertise used in sound design and musical instrument modeling. We also developed, based on the Perceptual Evaluation of Audio Quality (PEAQ) algorithm (ITU-R BS.1387), a robust methodology for objective prediction of sound synthesis quality.

An important feature of the novel framework is that it makes it possible to objectively predict sound synthesis quality in a automated synthesis parameters optimization setup which is useful for sound synthesis and evaluation work involving large sound databases. The framework can also be used for perceptual evaluation of the impact of individual synthesis parameters on the final sound quality. This has been found very valuable to benchmark sound synthesis methods.

Pipe organ sound design was used as a vehicle to illustrate the use of our system. However, the approach is generic and can be adapted for any sound analysis and synthesis methods for a number of musical instruments.

Results show that the approach is an important alternative to existing methods and that it can be used for other types of musical instruments, such as piano in Hamadicharef (2005). The new approach can serve as a basis for benchmarking sound synthesis and the development of instrument specific sound synthesis quality index.

Contributions and conflict of interest

Both authors have contributed equally to the research and approved the final article. The authors declare no conflict of interest related to this work, financial or otherwise.

Acknowledgment

The work reported here was supported by an EPSRC Research Grant (No. GR/S00859/01). We acknowledge support from the UK

Government Science Research Investment Fund for the purchase of the Opera *Voice/Audio Quality Analyzer* (Opticom GmbH, Germany). We are grateful to T. Koorlander (Musicom Ltd., Bideford, UK) and G. Blyth (UK) for sharing their expertise, support and fruitful discussions. We also would like to thank Dr. R. Clark (Allen & Heath Ltd., Penryn, UK) for his constructive suggestions.

References

- Ando, S., & Yamaguchi, K. (1993). Statistical study of spectral parameters in musical instruments tones. Journal of the Acoustic Society of America, 94, 37–45.
- Cheung, N.-M., & Horner, A. B. (1996). Group synthesis with genetic algorithms. Journal of the Audio Engineering Society, 44, 130–147.
- Comerford, P. J. (1993). Simulating an organ with additive synthesis. Computer Music Journal, 17, 55–65.
- Dolson, M. (1986). The phase vocoder: A tutorial. Computer Music Journal, 10, 14-27.
- Fletcher, N. H., & Rossing, T. D. (1998). The physics of musical instruments (2nd ed.). Springer-Verlag. ISBN:0-38798-374-0.
- Garibaldi, J. M., & Ifeachor, E. C. (1999). Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation. *IEEE Transactions on Fuzzy Systems*, 7, 72–84.
- Goldberg, D. E. (1989). Genetic algorithms in search. Optimization and machine learning. Addison-Wesley. ISBN:0-121-94270-8.
- Hamadicharef, B. (2005). Artificial Intelligence-based Approach to Modeling of Pipe Organs. Ph.D. Thesis School of Computing, Communications and Electronics (SoCCE), University of Plymouth Drake Circus, Plymouth PL4 8AA, Devon, United Kingdom.
- Hamadicharef, B., & Ifeachor, E. C. (1999). Artificial intelligence based modeling of musical instruments. In Proceedings of the 2nd COST-G6 workshop on digital audio effects (DAFx99), Trondheim, Norway, December 9–11, 1999 (pp. 117–120).
- Hamadicharef, B., & Ifeachor, E. C. (2003). Objective prediction of sound synthesis quality. In Proceedings of the 115th audio engineering society convention, New York, USA, October 10–13, 2003.
- Hopkins, E. J., & Rimbault, E. F. (2000). The Organ Its History and Construction. Bardon Entreprises Portsmouth. ISBN:0-952-81846-9.
- Horner, A. B., Beauchamp, J., & Haken, L. (1993). Methods for multiple wavetable synthesis of musical instrument tones. *Journal of the Audio Engineering Society*, 41, 336–356.
- Horner, A., Beauchamp, J., & So, R. (2006). A search for best error metrics to predict discrimination of original and spectrally altered musical instrument sounds. *Journal of the Audio Engineering Society*, 54, 140–156.
- ITU-R Recommendation BS.1115. (1993). Low Bit-rate Audio Coding. International Telecommunication Union, Geneva, Switzerland.

- ITU-R Recommendation BS.1116. (1997). Methods for the subjective assessment of small impairments in audio systems including multi-channel sound systems.
- ITU-R Recommendation BS.1387. (1998). Method for Objective Measurements of Perceived Audio Quality (PEAQ).
- ITU-R Recommendation BS.562-3. (1997). Subjective Assessment of Sound Quality. Keyhl, M., Schmidmer, C., & Wachter, H. (1999). A combined measurement tool for the objective perceptual based evaluation of the compressed speech and audio signals. In Proceedings of the 106th audio engineering society convention, Munich, Germany, May 8–11, 1999.
- Kleczkowski, P. (1989). Group additive synthesis. Computer Music Journal, 13, 12–20.
- Kostek, B., & Czyzewski, A. (2001). Representing musical instrument sounds for their automatic classification. *Journal of the Audio Engineering Society*, 49, 768–785.
- Laroche, J., & Meillier, J.-L. (1994). Multichannel excitation/filter modeling of percussive sounds with application to the piano. *IEEE Transactions on Speech and Audio Processing*, 2, 329–344.
- Lee, K., & Horner, A. B. (1999). Modeling piano tones with group synthesis. Journal of the Audio Engineering Society, 47, 101–111.
- Oates, S., & Eaglestone, B. M. (1997). Analytical methods for group additive synthesis. Computer Music Journal, 21, 21–40.
- Rioux, V. (2000). Methods for an objective and subjective description of starting transients of some flue organ pipes – Integrating the view of an organ builder. *Acta Acustica*, 86, 634–641.
- Sandell, G. J., & Martens, W. L. (1995). Perceptual evaluation of principalcomponent-based synthesis of musical timbres. *Journal of the Audio Engineering Society*, 43, 1013–1028.
- Smith, J. O. (1992). Physical modeling using digital waveguides. Computer Music Journal, 16, 74–87.
- Smith, J. O. (1996). Physical modeling synthesis update. Computer Music Journal, 20, 44–56.
- So, K. F., & Horner, A. B. (2002). Wavetable matching of inharmonic string tones. Journal of the Audio Engineering Society, 50, 46–56.
- Thiede, T., Treurniet, W. C., Bitto, R., Schmidmer, C., Sporer, T., Beerends, J. G., et al. (2000). PEAQ – The ITU standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Society*, 48, 3–29.
- Wang, Z., & Bovik, A. C. (2002). A universal image quality index. IEEE Signal Processing Letters, 9, 81-84.
- Wun, S., & Horner, A. (2005). A comparison between local search and genetic algorithm methods for wavetable matching. *Journal of the Audio Engineering Society*, 53, 314–325.
- Wun, S., & Horner, A. (2006). Evaluation of iterative methods for wavetable matching. Journal of the Audio Engineering Society, 53, 826–835.
- Zadeh, L. A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, *11*, 199–227.
- Zwicker, E., & Fastl, H. (1999). Psychoacoustics Facts and models (2nd ed.). Springer Verlag – Series in Information Sciences. ISBN:3-540-65063-6.