Artificial Intelligence based Modeling of Musical Instruments

Brahim Hamadicharef, Emmanuel Ifeachor

Audio Engineering Society

Convention Paper 5484

Presented at the 111th Convention
2001 September 21–24 New York, NY, USA

An intelligent system approach to sound synthesis parameter optimisation

Brahim Hamadicharef and Emmanuel Ifeachor
Department of Communication and Electronic Engineering
University of Plymouth
Plymouth, Devon PL4 8AA, England
brahim@cis.plym.ac.uk ; E.Ifeachor@plymouth.ac.uk

ABSTRACT

An intelligent audio system for sound design using artificial intelligence techniques is reported. The system is used to analyse acoustic recordings, extract salient sound features and to process them to generate parameters for sound synthesis, in a manner that mimics human audio experts. Preliminary tests show that the use of the system reduces design time and yet the quality of the resulting sound is considered high by audio experts

INTRODUCTION

A new approach to sound design and modeling of musical instruments is presented. An intelligent audio system, based on fuzzy logic techniques, is used to analyse acoustic recordings, extract salient sound features and to process them to generate parameters for sound synthesis, mimicking human audio experts.

The main goal of our research is to investigate and develop artificial intelligence based techniques to capture and exploit audio expertise in the design of high quality sound. Our principal aim is to automate, as far as possible, the complex and time-consuming task of sound design for musical instruments, by exploiting the experience and knowledge of professionals such as musical instrument manufacturers, audio engineers and musicians.

The project is being carried out in collaboration with two audio companies, one of which has expertise in organ pipe sound synthesis. To our knowledge, this is the first attempt in computer music to capture and exploit, explicitly, knowledge from audio experts for sound design.

In this paper a description of the concept and implementation of an intelligent audio system is given together with preliminary results.

SOUND DESIGN ENVIRONMENT

The sound design environment depicted in Figure 1. It consists of an acoustic unit, an intelligent audio system and an electronic unit.

The acoustic unit is an organ pipe (other acoustic musical instruments may be used, but we are using organ pipe as a vehicle for the project). The sound of the organ pipe is recorded using microphones placed at different positions along the pipe. The intelligent audio system is a dedicated multimedia computer with a sound card and a MIDI a interface. The computer is also connected to the sound generator to transfer the setup parameters generated by the intelligent system. The electronic unit is based on a sound generator, supplied by our collaborative companies, connected to a MIDI master keyboard also controlling the multimedia computer via MIDI.

A digital mixer interconnects to all three units of the system. The audio expert or user can experiment within the sound design environment, e.g. play the acoustic instrument, record the sound onto disk, use the intelligent audio system to design sounds and listen to synthesized sound to assess their quality.

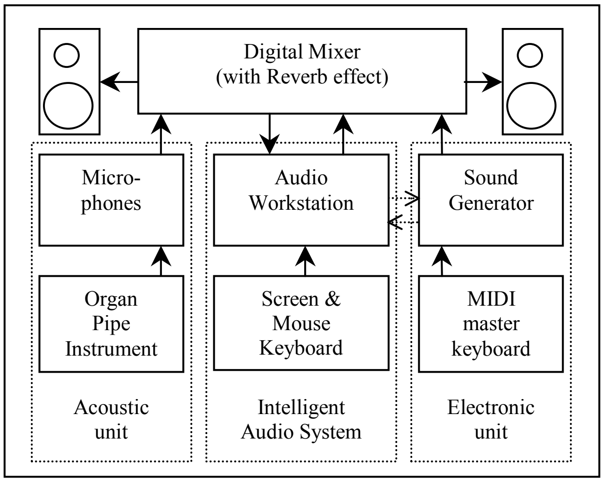

Figure 1: Diagram of the sound design environment.

The digital mixer may be used to add some reverberation to the synthetic sound to re-create a church acoustic environment typical to organ pipes.

INTELLIGENT AUDIO SYSTEM

A conceptual diagram of the intelligent audio system is shown in Figure 2. The intelligent audio system is implemented as software tools written within the MATLAB environment and linked to other research tools and custom applications written in C++. The intelligent audio system is divided into three main parts: a sound analysis engine, an intelligent system based on audio expertise and a sound synthesis engine.

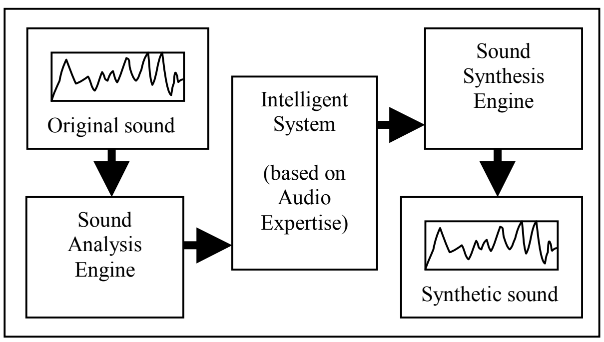

Fig. 2. Conceptual diagram of the intelligent audio system.

Sound analysis engine

The sound analysis engine serves as a front-end of the intelligent audio system. A block diagram of the sound analysis engine is shown in Figure 3. Using our sound design environment the audio expert or user can record organ pipe sounds and save them onto disk. In developing the intelligent audio system we have used a bank of sound from a CDROM. The sounds are standard WAV files, with two channels, the first being a recording very close to the mouth of the organ pipe and the other one near the extremity of the pipe. The user of the intelligent audio system loads a sound file and runs the sound analysis engine. For certain sounds we have found that some pre-processing is required to normalise the sound and remove background noise due to the recording conditions. Amplitude, frequency and phase trajectories are computed using phase vocoder-based techniques [1][2]. Examples of temporal and time–varying spectral representations of an organ pipe sound are given in Figure 4 and Figure 5 respectively.

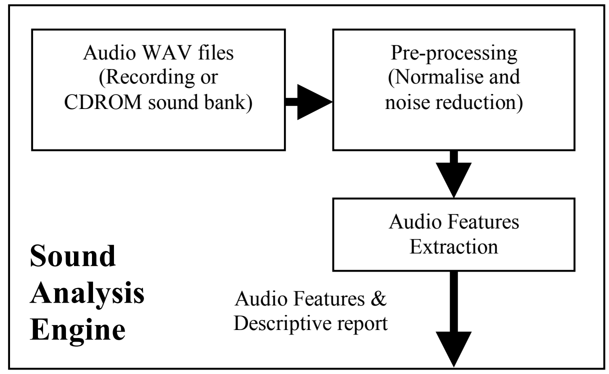

Figure 3: Block diagram of the sound analysis engine.

Audio features

The analysis engine further extracts audio features from time-varying harmonic components of the sound. The audio features can be divided into two categories: temporal and spectral features.
Temporal features correspond to the time evolution of the overall amplitude envelope of the sound. The amplitude envelope has typically five segments: start, attack, sustain, release and end, with sometimes an additional decay segment between the attack and sustain. Spectral features related to time-varying evolution of the spectral envelopes.

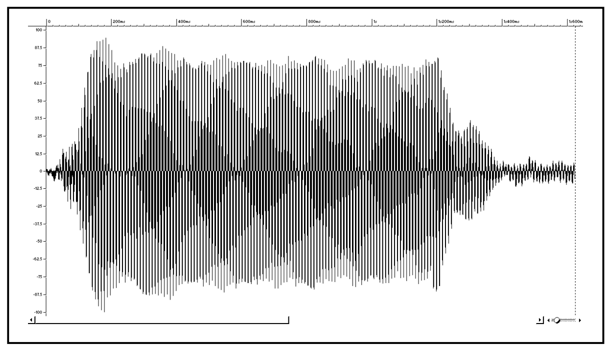

Figure 4: Temporal representation of an organ pipe sound.

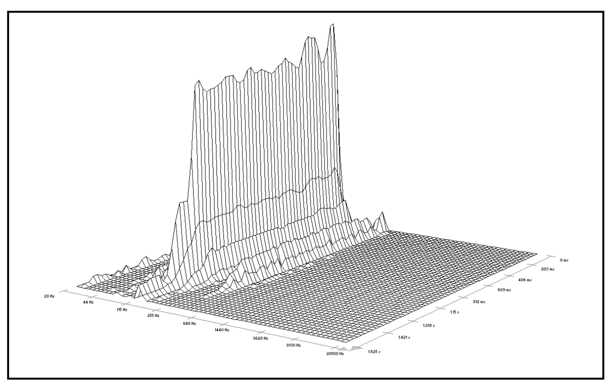

Figure 5: Time-varying spectral representation of an organ pipe sound.

Audio features extraction

In the following section methods used to extract the audio features will be detailed.

To extraction the temporal features of sound we model the amplitude envelope using a split-point time estimation technique [3]. This involves smoothing the envelope by convolution (digital signal processing technique) of the envelope with a gaussian. Using time derivatives of the smoothed envelope, the location of the start – end of each of the portion of the envelope can be determined (i.e. Attack, Decay, Sustain, Release). Each split-point has a variable amplitude (in percentage of the maximum amplitude of the partial) and time. The shape of each segment can be exponential, linear or logarithmic. This method has proved to be more stable than conventional methods.

The features extraction process focuses on 7 main spectral features: the energy of the fundamental, the energy of harmonics 2 to 4, the energy of the remaining harmonics, the frequency of the fundamental, the spectral centroid, the ratio of the energy of even harmonics to the total energy of the signal, and the ratio of the energy of odd harmonics to the total energy of the signal excluding fundamental [4][5].

Descriptive Report

A descriptive report is generated to characterise the general impression, the transient and steady state parts of the organ pipe sound. The linguistic terms are those used by audio experts to describe organ pipe sound. Similar work has recently been published in [6] (see table 2).

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

Table 2: Linguistic descriptors used by the expert for the description of organ pipe.

Intelligent System

The intelligent system is based on fuzzy logic concepts [7][8] and is used to process audio features using rules provided by the audio experts to generate suitable parameters for the sound synthesis engine. A diagram of the fuzzy intelligent system engine is shown in Figure 6. First the fuzzification process transforms the extracted audio features into fuzzy variables. The fuzzy inference engine applies the rules to the fuzzy variables. The rules determine how the features are clustered / grouped based on the experience of audio experts. Finally, the output of the fuzzy inference engine are defuzzified and used to generate synthesis parameters to configure and control the sound synthesis engine.

Sound synthesis engine

The sound synthesis engine is currently based on a software tool that implements multiple wavetable synthesis with advanced modulation. It generates a sound file (WAV file) using the parameters from the intelligent system. A diagram of the sound synthesis engine is shown in Figure 7.

The audio expert can listen to the synthetic sound from the computer connected to the digital mixer and speakers (or headphones not shown in Figure 1). Listening and comparing the synthetic sound with the original allows the audio expert to assess the synthetic sound quality and hence the system performance.

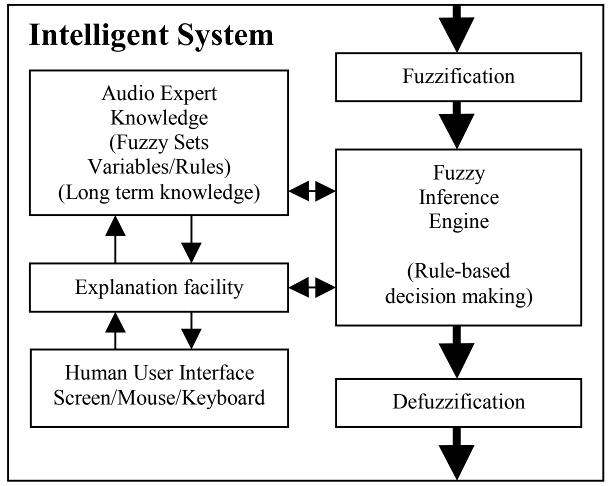

Figure 6: Diagram of the intelligent system.

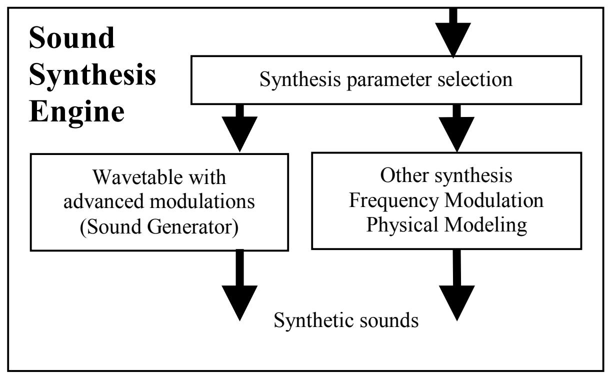

Figure 7: Diagram of the sound synthesis engine.

TESTS

To validate the new approach, preliminary tests have been conducted with a large bank of organ pipe sounds provided by one of the collaborative companies. It includes sounds from the Pedal (for the lower pitches that are played on the pedal board of the organ), Great, Swell and Choir departments of the organ pipe keyboard. Listening tests were conducted between the original sound and the sound generated from our intelligent audio system.

RESULTS AND PERFORMANCE

The preliminary results show that:

- The quality of the sounds generated by our system is often indistinguishable from the original one.
- The use of the system reduced the time taken to design organ pipe sounds to a quality that is considered high by two audio experts.

FUTURE WORK AND CONCLUSIONS

In future, the intelligent audio system will be extended in several ways. First we will to extend it to generate parameters for a complex, hardware-based, real time additive synthesis system which is of interest to the collaborative companies. The quality of the sound synthesis will be evaluated objectively using perceptual-based methods. At present, only subjective methods are used to assess the performance of synthesis systems of musical instruments. The rules used in the fuzzy expert system will to be refined to cater for sounds considered as more challenging by audio experts. Further, we will cater for other types of sound synthesis techniques such as Frequency Modulation [9][10] and

physical modeling (e.g. digital waveguides) [11][12]. Finally, we will investigate ways of extending our techniques to include other types of instruments such as bell [13] and piano [14].
An intelligent audio system for sound for organ pipes is reported. The intelligent audio system mimics human audio experts. To our knowledge, this is the first attempt in computer music to capture and exploit, explicitly, knowledge from audio experts for sound design.

ACKNOWLEDGEMENTS
The authors wish to acknowledge the support and assistance of Tony Koorlander and Graham Blyth of Musicom Ltd. We also would like to thank Rob Clark of Allen & Heath Ltd for his constructive suggestions and help.

References

[1] De Gotzen, A., Bernardini, N., and Arfib, D., 2000. "Traditional implementation of a phase vocoder: the tricks of the trade", Proceedings Workshop on Digital Audio Effects (DAFx-00), Verona, Italy,
[2] Laroche, J. and Dolson, M., 1999 "Improved phase vocoder time-scaled modification of audio". IEEE Transactions on Speech and Audio Processing. Vol. 7, No. 3 May, pp 323-332
[3] Jensen, F., 1999. "Envelope Model for isolated musical sounds", Proceedings Workshop on Digital Audio Effects (DAFx-99), Throndeim, Norway, 1999.
[4] Kostek, B., 1995. "Statistical versus artificial intelligence based processing of subjective test results" Proceedings of the 99th Audio Engineering Convention Preprint 4018 (P-3), Paris, France, February 25-28, 1995.
[5] Kostek, B., 1995. "Feature extraction methods for the intelligent processing of musical instruments" Proceedings of the 99th Audio Engineering Convention Preprint 4076 (H-4), New York, USA, October 6-9, 1995.
[6] Rioux and Vastfjall, 2001. "Analyses of verbal descriptions of the sound of the flue organ pipe". Musicae Sciencae, Volume 5, Number 1, Spring.
[7] Zahed, L. A., 1983. "The role of fuzzy logic in the management of uncertainty in expert systems", Fuzzy Sets and Systems, Vol. 11, pp. 199-227.
[8] Cox, E. The fuzzy system handbook, AP Professional
[9] Horner, A., 1998 "Nested Modulator and Feedback FM matching of Instrument Tones", IEEE Transactions on Speech and Audio Processing. Vol. 6, No. 4, July.
[10] Tan, B. T. G., and Lim, S. M., 1996. "Automated parameter optimisation for double frequency modulation synthesis using the genetic annealing algorithm". Journal of the audio Engineering Society, Vol. 44, No. 1/ 2, January/February 1996, pp. 3-15
[11] Smith, J. O., ``Physical modeling using digital waveguides'', Computer Music Journal, Vol. 16, No. 4, pp. 74-91, Winter 1992
[12] Valimaki "An improved digital waveguide model of a flute with fractional delay filters", 96th Nordic Acoustic Meetings, Helsinki, 12-14 June 1996
[13] Horner, A., Ayers, L. and Daniel Law, D. "Modeling Small Chinese and Tibetan Bells" Journal of the Audio Engineering Society, Vol. 45, No. 3, 1997 March
[14] Zheng, H., and Beauchamp, J., "Spectral characteristics and efficient critical-band-associated group synthesis of piano tones" Journal of the Acoustic Society of America, Vol. 106, No. 4, Pt. 2, pp. 2141-2142.