DESIGN AND EVALUATION OF PHYSICALLY INSPIRED MODELS OF SOUND EFFECTS IN COMPUTER GAMES

Böttcher, Niels; Serafin, Stefania

Published in:
AES 35th International Conference, London, UK, 2009 February 11–13

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

DESIGN AND EVALUATION OF PHYSICALLY INSPIRED MODELS OF SOUND EFFECTS IN COMPUTER GAMES

NIELS BÖTTCHER[1], STEFANIA SERAFIN[2]

[1] *Institute 22, Medialogy interaction and sound group , Aalborg University, Ballerup, Denmark*
nib@imi.aau.dk
[2] *Institute 22, Medialogy interaction and sound group , Aalborg University, Ballerup, Denmark*
sts@imi.aau.dk

A simple audio only game was designed in order to test the possibilities of different interactive sound synthesis techniques simulating aerodynamic sword-like sounds. The Wii remote was used as the controller for generating the sound and this device was connected to the Max/MSP sound synthesis engine. In order to gain a user centered perspective on the potential of the diverse synthesis techniques, in terms of sound quality, interactivity and entertainment value, several techniques to synthesize real time interactive sword-like sounds were implemented. In this test a sample based model was compared to physically inspired modal synthesis, purely perceptually modeled subtractive synthesis and granular synthesis.

INTRODUCTION

Sound effects in computer games are currently developed by using sampled sounds which are manipulated in different random ways in order to cope with repetitions. The main audio features that can be manipulated or controlled in current game engines are sample playback, volume and spatial properties of the sound source. Most game companies tend to focus on multilayered pre-recorded soundscapes and not much effort is put on interactive action or event sounds. This is most likely due to a choice of emphasizing the graphical elements in the games and in order to use most of the computational power on the graphic rendering .

With novel gestural control devices like the Nintendo Wii, that gives the player a much higher level of variety in terms of interaction, we believe that there will be a higher demand for interactive sound synthesis in computer games .

An alternative to the use of sample based audio in computer games, is the development of sound synthesis algorithms based on physical models which simulate a particular sound event. The obvious advantage of using physical models is the fact that the game designer is able to manipulate all the different physical parameters of the model according to the actions of the game. This could be parameters like velocity of the excitation as well as material of the objects among other things.

However, one current drawback of interactive physical models, which limits their use in commercial game engines, is the fact that the sound quality is not satisfactory enough as the one of recorded samples.

We believe that physically inspired sound models, together with appropriate controllers, have a strong potential in computer games.

By physically inspired sound models we mean synthetic sounds whose creation is originated from physics phenomena, but whose implementation has more focus on the way the sounds are perceived by the end user rather than on an accurate physical simulation of the phenomenon itself.

1 RELATED WORK

Interactive sound synthesis is not a new topic in relation to computer games. Andy Farnell[1][2] has done a great deal of work on procedural audio and interactive sound synthesis using several different synthesis techniques. Among others the examples counts procedural synthetic footsteps by the use of granular synthesis[3].

Other examples of interactive footsteps include[4] where modal synthesis is used to synthesize sound created by walking on different materials.

In[5] a real-time interactive physical model of aerodynamic sound is presented and implemented for sword sounds.

Several techniques for applying interactive sound synthesis in interactive applications, is presented in [6] by Perry Cook. This includes different physical models as well as spectral modelling and additive synthesis techniques.

In [7] techniques for using granular synthesis for creating sound effects and ambiences in computer games is proposed.

Environmental sounds, collision sounds and similar can also be synthesized in real-time. In [8] techniques for creating more efficient and realistic sound simulations by physics in computer games are presented.

This paper is presenting the early work of a PhD thesis about interactive action sounds for computer games

with novel controllers. The emphasis is on sword-like sounds and the following presents the implementation and evaluation of 4 different interactive models synthesizing sword-like sounds.
The Wii remote was used to control the sound synthesis as the control device represents a novel and very popular control device for contemporary computer games. We believe that control devices similar to the Wii remote is here to stay and it is only a matter of time before we will see some even more sophisticated controllers on the commercial marked.

2 IMPLEMENTATION

2.1 Connecting the Wii to Max/MSP

In order to connect the Wii remote to Max/MSP the OSCulator software[9] was used. The Wii remote was connected to the OSCulator software and the data was retrieved through the open Sound Control protocol. We used the Wii remote OSCulator 2.5 pre-made Max patch developed by kim cascone[10] in order to handle the data sent from OSCulator and the Wii remote.

2.2 Sample sounds

High quality samples of sword sounds were recorded in a professional recording studio using a high end Neumann U87 Microphone and a Danfield MX10 preamp.
In order to achieve the best sounding sword-like sound a lot of different swords and sticks of different materials were tested. To our surprise the best and most promising sword sound came from a thin 1 meter long wooden bamboo stick.

In Max/MSP two different sword samples were implemented. One aggressive sword strike and one more gentle.
A threshold was applied in order to set a level of when the strike was hard and when it was more gentle.
The user could now achieve the playback of two different sounds depending on the acceleration of the Wii remote.
The acceleration was furthermore mapped to the level of amplitude of the samples.
It was only possible to play one sample at each strike with the sword.
The purpose of the sample patch was implement a patch similar to the current status of action sounds in commercial and common game engines.

2.3 Subtractive synthesis

This model was based on filtered noise. A low amplified continuous noise signal was sent through a bandpass filter. The acceleration of the Wii remote was mapped to the cut-off center frequency of the bandpass filter, moving between 100Hz and 3500Hz in a linear mapping.
Additionally the bandwidth of the filter was mapped to the acceleration of the Wii remote as well. A high threshold was set in order to only perform a preceptive change in the slope of the filter, when the acceleration of the Wii remote was significantly high.
The estimation of the frequency scaling as well as the regulation of the slope of the filter, was made purely of how the author perceived this as the most sword-like output.

2.4 Granular synthesis

For the granular synthesis the pre-made Granulized patch[11] from the Max/MSP 5 examples was used.
The same sword sample as the aggressive sample in the sample model was used as the basis of the granulation.
The duration of the grains where set as low as possible and the spacing between the grains were set to 20 ms. The pitch of the sound was slightly lowered in order to speed up the sample during the acceleration of the Wii remote.
The acceleration of the Wii remote was mapped to the scrubbing through the different grains of the sample. The Amount of acceleration was determining the playback speed of the scrubbing as well as how far in the sample the playback would go.
The original sample of the 1 meter long wooden bamboo stick had a high pitch swoosh sound in the end of the sample. If the acceleration was below a certain threshold this part of the sample would never be played.
Additionally the acceleration was also mapped to the amplitude of the overall granular synthesizer.

2.5 Physically inspired model

The physically inspired model was based on modal synthesis, where you divide the model in an exciter and a resonator.
As the excitation of the model, a filtered sample was used. The recording of the wooden bamboo stick from the sample patch was analysed and the most significant resonant peaks where filtered out.
In this case 5 significant areas of resonant peaks where found and filtered out. This was done with a band reject filter.
As the resonator for the model a simple one-zero, two-pole bandpass filter was used. The bandpass filter had the filtered sample as input and the 5 resonant peaks were now replaced by using 5 bandpass filters in parallel emphasizing the resonant peaks.
The bandwidth of the different resonant peaks were all set to the same value.
The lower frequencies of the filter was set with a slightly higher amplitude as this was found in the spectral analysis of the samples.

The acceleration of the Wii remote was mapped to the amplitude of the different resonant peaks as well as scaling the frequencies slightly within a specified level. This was also found in the analysis of the sample, that the resonant peaks were slightly changed according to the acceleration of the wooden stick. The mapping of the acceleration was made as an estimation..
Furthermore the overall amplitude of the model was mapped to the acceleration of the Wii remote.

2.6 The panning of the sound

At the first implementation of the models the infrared camera in the Wii remote was used in order to track the spatial localization of the remote. It was possible to track the spatial location within a specified area between the infrared light sources. Unfortunately this area seemed to be to limited and unstable in order to use it for the panning of the sound.
Instead the horizontal panning of the sound was mapped to the horizontal orientation of the Wii remote.
This seemed to work satisfactory, when performing body gestures similar to having a sword in the hands.

2.7 The audio only game

In order to make the test persons have a reason for playing with the Wii controller, a simple audio-only game was implemented.
The game was kept very simple and the purpose of the game was mainly to get the players to utilize the possibilities of the of the Wii remote as well as the mapped interactive sounds.
The basic narrative in the game is that you are placed blindfolded in front of the most irritating guy in the world. You have a sword-like thing in your hands and what can you do? You simply have to punish the guy!
But as you are blindfolded you can't see the guy and you have to strike blindly around yourself until you hit the guy. In the beginning of the game you will hear the voice of the opponent and he will also be located at a random place around the horizontal plane. But very soon he is gone and you can't hear him anymore. Now you simply just have to hit all around the place and suddenly you will strike the guy.
When the player finally hits the guy, a collision-impact sound is played followed by a wail from the opponent.

2.8 Background sounds

In order to test the interactive sound in relation to a realistic game environment, a background soundscape was implemented.
When the players pushed the start button a background soundscape was initialized. This was to indicate that the game was now running. The mood of the background was kept in a puzzling mysterious way in order to give some suspense to the game.
Shortly after the activation of the game a voice of the opponent is played and this should indicate that you could start hitting after the guy.
During the game randomly panned sounds of footsteps and crackling branches are played. This was done in order to give the player a hint of where the opponent was hiding, and in order to keep the suspense of the game.

3 TEST

In order to gain a user centered perspective of the potential as well as limitations of the different synthesis techniques a test was performed.

3.1 Setup of the test

The test was performed in a quiet office environment, where the test persons were placed in front of a set of M-Audio StudioPhile stereo monitors, with the Wii remote in their hands. The test subjects were asked to close their eyes during game.
Before starting the test, the subjects were asked to try out the application and play around with the game for a approximately 1 minute. The model for this task was chosen randomly before each test.
When the test subjects had been playing around with the Wii remote and seemed as if they had a good understanding of what the game was about, another model was opened. The test persons were asked to try out the game and then fill out a questionnaire of 5 questions related to the specific sound synthesis they had just been interacting with.
The same procedure was repeated for the last 3 models afterwards.
The sequence of the models were chosen randomly for each test subject.
When the test persons had tried out all the different models, a few conclusive questions were asked, in order to compare the different models.
Eventual comments or interesting observations were noted down during the test.
The test lasted for approximately 20-25 minutes.

3.2 Test persons

The test was made on 15 test subjects in the age of 20-47 years. The average age of the subjects were 29 years. All of the test subjects were students or teachers at the institute 22 at Aalborg University Copenhagen.
It was a requirement that all the test persons had a basic prior knowledge with a Wii controller.

3.3 The test questions

The test subjects were given a questionnaire consisting of 5 questions for each of the different model. They were asked to write down what they thought the sound was simulating first of all.

In the following the subjects were asked to rate the following:

- The sound quality
- How realistic the sound was
- How the interaction was
- How the connection with the gestures was

The rating was done in a scale from 1-5 and for all the answers, it was possible to write down eventual comments.

At the end of the test, the subjects were given a 4 concluding questions in relation to the following:

- If they preferred any of the models more than the others
- If there was a difference in the nuances of the interaction for any of the models
- If there was any of the models that seemed to be more entertaining than others

1 RESULTS

1.1 How did the users perceive the sound quality of the different models?

In general the average test person judged the granular synthesis patch as having the best sound quality.
The mean value of the rating among all the test subjects was set to above average of the sound quality for this patch.
Very close to this was the subtractive synthesis model that was also rated a little bit above average.
The sample model was judged to be just above middle and the only patch where the sound quality was judged to be below average was the modal synthesis patch.

1.2 How realistic did the users find the different models?

Again the granular synthesis model seemed to be judged as the most realistic patch and very close after this model came the patch based on samples.
The subtractive synthesis model as well as the physically inspired modal synthesis patch was judged to be below average.
The realism was judged in comparison to everyday sounds and not in relation to the interaction or in relation the the performed body gestures.

1.3 How was the connection between the gesture and the sound?

Regarding the connection between the gesture and the sound produced, the granular synthesis as well as the subtractive synthesis, was rated as above average. The patch using samples was rated below average and surprisingly also the physically inspired modal synthesis was rated as being below average.
It seemed as if the sound quality was also influencing the way people perceived the level of interaction and connection with their body gestures.
It was noticed during the observations of the test persons, that the nuances and variations in body gestures was very limited with especially the patch based on samples. People started adapting to the limitations of the patch and their movements became much more monotonous.
This was exactly opposite in regards to the subtractive synthesis patch. The subjects were tending to change their body gestures into much more varied and detailed movements. One of the test subjects even mentioned that “you feel like doing a lot of Zorro-ish moves”.
Regarding the modal synthesis patch, the test subjects tended to be irritated of the sound quality and they seemed to quit this test before some of the other sound models.

1.4 Which model did the users prefer?

For the question regarding what model the subjects would prefer and found the best, it was very clear that the granular synthesis patch did significantly better than all the other models. 57.1 % of the user rated this model as the best of the 4 different models.

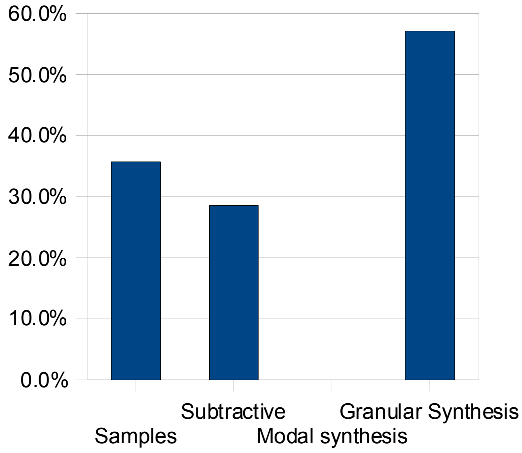

Figure 1: 57.1 % of the test subjects preferred the granular synthesis patch more than the others.

The patch using sampling was rated the best by 35.7% of the users and the subtractive synthesis patch got 28.6% of the votes.
The physically inspired modal synthesis patch was not voted as the best solution for any of the users at all.
The arguments for choosing the granular synthesis model over the others was among others things, that it seemed more realistic, that there was a lot of variance in the sound, and that it was more fun to play with.

1.5 Was there a difference in the nuance of interaction?

Very surprisingly there was a rather large percentage of the test persons that did not find a significant difference in the nuances of the interaction with any models more than others. 42.9 % of the subjects answered no to this question, after having utilized all of the patches.

By observing how the users were utilizing and playing with the Wii remote during the game, there was a strong difference in the body gestures, depending on what sound that was mapped to the controller.
When playing with the model using sampled sounds, the users had a tendency of making the same gestures again and again, and they did not spend as much time on this model compared to the others.
The subtractive synthesis patch was very interesting in relation to this, as this sound made the users move in much more detailed ways and with much more diversity.
This was one of the the most interesting observations of the test, and it was rather surprisingly how significant the different gestural behaviours was related to the level of interaction possibilities and the corresponding sound.

1.6 Was there any of the models that seemed more entertaining than others?

For this question the answers were well spread between the granular synthesis, the sample patch and the subtractive synthesis patch. 35.7 % of the test subjects rated the subtractive synthesis patch as the most entertaining patch.
Both the sampling patch as well as the granular synthesis patch had 28.6 % of the votes for the most entertaining patch.
None of the test persons rated the physically inspired modal synthesis patch as the most entertaining.

When asked why they chose the subtractive synthesis model as the most entertaining, the answers were highly related to the interaction with the sound.
Some of the subjects thought this model was responding well to the movements and other mentioned that they performed different movements when playing with this sound.
This also seemed very clear through the observation of the test subjects as mentioned earlier in the paper.

1.7 Additional observations and things to improve for the next test

It was very clear from the test that the sample based model was not engaging for the test persons, in terms of the interaction with the models, and mapping between gesture and sound.
To our surprise the physically based modal synthesis patch was not doing very well in any of the tests. This was the case in relation to the sound quality but also in terms of interaction and connection between the gesture and sound.
The implementation of the physically inspired model was based on a sampled sound as the residual input. This gave some limitations in terms of the interaction.

It was clear that the patch based on subtractive synthesis with a bandpass filter and pure continuous noise as input, was rated much higher in all the different cases.
It would be interesting to test the physically inspired models on a continuous input instead of a limited sample of the residual.
Also the sound quality of the modal synthesis needs to be improved in order to test the interaction possibilities with the sound. This seemed to have a large impact on the perception of interaction.

2 CONCLUSIONS

This paper describes how a simple audio-only game was designed in order to test the possibilities and limitations of different interactive sound synthesis techniques simulating aerodynamic sword-like sounds. The Wii remote was used as the controller for generating the sound and this device was connected to the Max/MSP sound synthesis engine.
Four different sound synthesis techniques was implemented, using subtractive synthesis, granular synthesis, samples and physically inspired modal synthesis.

In order to compare the qualities and limitations of the different models in terms of sound quality, interaction and entertaining level, a test was performed.
The test showed a significantly higher rating in terms of sound quality as well as interaction in the case of granular synthesis. This model was judged as the most successful implementation.
The subtractive synthesis patch, that was based on continuous filtered noise, was rated as the most entertaining of all the patches. With this patch it was possible to do more subtle and varied body gestures and it was also observed that most of the test subjects were changing their behaviour and body gestures significantly when playing with this sound synthesis mapped to the Wii controller.
As expected, the patch purely based on samples was rated as average or above average in terms of sound quality of the single strike, but in terms of entertainment, interaction and connection to the body gestures, this model was rated below average. The test subjects tended to become annoyed by the limitations and lack of variations when playing with this patch.

As a surprise the physically based modal synthesis patch was also doing very unsatisfactory both in terms of sound quality, but also in terms of interaction and entertainment level.
It seemed as if the sound quality had a rather impressive impact on the way the test persons would rate both the interaction, as well as the entertainment level of the models.
The physically inspired modal synthesis patch was based on a filtered sample as excitation. For future implementations of the model, it would be an idea to base this patch on a continuous input as excitation, in order to make the model more sensitive to continuous and more subtle body gestures.

The subtractive synthesis model which was based on the authors subjective perception of a sword sound only, was rated quite high. This was the case both in terms of interaction, but also in terms of sound quality.
This implementation seemed to be the most entertaining patch of different ones tested.

3 FUTURE WORK

It is the plan to continue the work of the implementation of the physically inspired models. It is very clear that the quality of the sound needs to be improved and that more aspects from the spectral analysis of the sound has to be considered.
A continuous input seems to be very important when interacting with the sound. This input could for instance be a simple continuous pure noise input as used in PHISM [12] by Perry Cook.

An interesting point of view is also to look at the importance of the physical correctness of the spectral modelling. The subtractive synthesis patch that was based purely on how we subjectively perceived the sound. This model seemed to be much more successful among the test persons. It would be an interesting issue to test how cartoonish the users would prefer the sound in comparison to how physically correct.

Currently a test is being prepared, in order to see if there is a difference in how the users will perceive the nuances and possibilities of the interactive sound, when having a visual feedback.
This test will be performed with a simplified computer game implemented in the Unity game[13] engine in connection with the Max/MSP sound engine.

4 REFERENCES

[1] http://Obiwannabe.co.uk

[2] Andy Farnell, "An introduction to procedural audio and its applications in computer games" *Audio Mostly Conference* 2007,Ilmenau, Germany

[3] Andy Farnell, "MARCHING ONWARDS Procedural synthetic footsteps for video games and animation" *PD convention* 2007,Montreal, Quebec, Canada

[4] Rolf Nordahl, "Self-induced footstep sounds in virtual reality: Latency, recognition, quality and presence", in proceedings of the 8th annual Workshop on Presence, London, UK, 2005

[5] Yoshinori Dobashi et. al., "real-time rendering of Aerodynamic Sound using sound textures based on computational fluid dynamics", ACM transactions of graphics (TOG), volume 22, issue 3, juli 2003

[6] Perry R Cook, "Real sound synthesis for interactive applications", A.K. Peters Press, 2002

[7] Leonard J. Paul, "An introduction to granular synthesis in video games", From pacman to pop music – interactive audio in games and new media, edited by Karen Collins, Ashgate publishing 2008

[8] Nikunj Raghuvanshi et al., "Real-time sound synthesis and propagation for games",Communication of the ACM, Volume 50, issue 7, juli 2007

[9] http://www.osculator.net/wiki/

[10] http://anechoicmedia.com/

[11] http://www.cycling74.com/products/max5

[12] Perry Cook, "Physically informed Sonic Modelling (PhiSM): Synthesis of percussive sounds", Computer Music journal 21(3): 38-49 (1997)

[13] http://unity3d.com/