Evaluation of Car Engine Sound Design Methods in Video Games

Harald af Malmborg

Audio Technology, bachelor's level
2021

Luleå University of Technology
Department of Social Sciences, Technology and Arts
Abstract

Realistic interactions with sound objects in video games are a contributing factor to the overall immersion. Car engine sound design is an area where the auditory feedback from driving would have an impact on that immersion. For this study, three methods for designing car engine sounds are evaluated. A traditional method in the game sound industry is to use several sampled audio recordings for the reproduction of engine sounds, which is in this study represented as a sample-based model. The sample-based model is evaluated together with a model using granular synthesis in an in-game scenario. A less common method is to use physical modeling, which is in this study evaluated together with the other two models in a listening test. Results show the granular synthesis model to be the most realistic, and the physical model to be the least preferred.
# Table of Contents

1. Introduction .................................................................................................................. 4  
   1.1 Procedural Audio ........................................................................................................ 4  
   1.2 Sound Synthesis Techniques .................................................................................... 6  
   1.3 Physical Modeling ...................................................................................................... 6  
   1.4 Granular Synthesis .................................................................................................... 7  
      1.4.1 Large Scale Granular Synthesis ............................................................................. 7  
      1.4.2 Concatenative Synthesis ....................................................................................... 7  
   1.5 Perceptual Evaluation of Synthesis Techniques .......................................................... 8  
   1.6 Auditory feedback from driving ................................................................................ 8  
2. Research Question, Aim, and Purpose ......................................................................... 9  
   2.1 Expected results ........................................................................................................ 10  
3. Method ............................................................................................................................ 10  
   3.1 Demographics ........................................................................................................... 10  
   3.2 Pre-Study .................................................................................................................. 11  
   3.3 Experiment design ..................................................................................................... 12  
      3.3.1 Turbomania .......................................................................................................... 12  
   3.4 Choice of Method ...................................................................................................... 14  
      3.4.1 Blend Container .................................................................................................... 15  
      3.4.2 REV .................................................................................................................... 15  
   3.5 Additional Sounds .................................................................................................... 16  
   3.6 Tools .......................................................................................................................... 16  
   3.7 Porsche 997 and 991 ................................................................................................. 17  
   3.8 Main Study ............................................................................................................... 17  
      3.8.1 Main Study Survey Questions .............................................................................. 17  
      3.8.2 Post-Study Listening Test Survey Questions ....................................................... 18  
   3.9 Post-Study Listening Test .......................................................................................... 19  
4. Results and Analysis ..................................................................................................... 21  
   4.1 Main Study Survey Responses ................................................................................... 21  
   4.2 Main Study Analysis ................................................................................................. 24  
   4.3 Post-Study Survey Responses .................................................................................... 25  
   4.4 Post-Study Analysis .................................................................................................. 27  
   4.5 Demographic data analysis ...................................................................................... 27  
5. Discussion ....................................................................................................................... 28
5.1 Critique of the Method
5.2 Pure Data
  5.2.1 The Heavy Compiler
5.3 Future Research
5.4 Conclusion
6. Bibliography

Appendix A: Full answers regarding if the driving experience felt better in one of the cars and why so.

Appendix B: Full answers from the post-study listening test survey regarding if any of the sounds would offer a better driving experience according to the subject’s preference.

Appendix C: Full experiment instructions for subjects.
1. Introduction

Our relationship to sound may change when we have some choices about it, for example when we can interactively alter the sound by shaping, creating, or selecting it. Game sound change depending on the player's action. If the player notices the connections between play and the sound the game makes, it can contribute to immersion. For the connection to have an immersive impact on the player, the sound needs to be believable to avoid breaking the immersion. Many game developers strive for believable interactions and therefore also believable sound experiences. The act of driving vehicles in games is a type of interaction where the player often expects believable responsiveness from the vehicle. Vehicle sound design in video games has traditionally revolved around sample-based methods. The specific vehicle for use has been recorded, sampled, and implemented into the game, with the sample changing when accelerating and shifting terrain. Although this approach could be seen as interactive to some degree, some methods could be considered more sophisticated and more interactive when designing sound for vehicles in video games. Procedural audio is key to many of these methods.

1.1 Procedural Audio

Farnell (2007) defines the term procedural audio as "... non-linear, often synthetic sound, created in real-time according to a set of programmatic rules and live input" (p. 1).

The non-linear aspect of game audio, contrary to the linear flow of events in film and television, is a challenge the sound designer has to face. Games can last over 100 hours, and during that time, some sound effects may be heard many times. After a while, they sound repetitive. To avoid this, historically, designers have included a set of varied samples. But not all sounds need to be varied. Some non-diegetic sound effects, sounds not linked to an event in the fictional world, such as menu sounds, heads up display (HUD) sounds, or collectible items as some examples can repeat. In contrast, any type of diegetic sound that is likely to appear frequently in a short period would most likely be perceived as repetitive if it was repeated without modification (Vachon, 2009).

When designing vehicle sounds for video games, a sample-based approach is traditionally used. Recordings are typically looped, pitch-shifted, and cross-faded according to information from the game, such as revolutions per minute (RPM), throttle load, and gear. Baldan et al. (2015) suggest the results, in general, are realistic and the method is relatively simple to implement,
however, there are a few drawbacks. Mainly, it requires huge banks of samples, as well as the timbre of the recorded material heavily constrain the sound designer's space for creative design. A more procedural audio approach could instead be suggested when designing sound for a car engine. Instead of utilizing a discrete event, the procedural approach would rely much more on real-time parameters, modifying these events separately.

In sound design, whether using either sample-based or procedural approaches, complex sounds, like car engines are constructed out of components. The components might be the ignition, a rev up, acceleration, gear shift, etc. The physical models for the sounds inside the combustion engine and exhaust are standing waves, the frequencies of which are determined by the engine revs, the length, and the construction of the exhaust. Variables like engine revs can be mapped to user input. Component construction could be beneficial for a couple of reasons, not only for car engines but when designing sound for other vehicles as well. Farnell (2007) argues, that according to psychoacoustic models of perception and Gabor's granular theory, it is possible to replace distant, less focused sounds with noise approximations. In largely populated game worlds, this can lead to cleaner mixes without the excess of sampled audio channels. According to Farnell (2007), the magic cutoff point where procedural sound outperforms samples is when the density of sound sources reaches a few hundred. The technique of component construction could be applied to the sound of a helicopter for example. Instead of simply using filters to attenuate the sampled helicopter sound according to a distance parameter, the sound can be separated into different components, and thereby create a much more complex and sophisticated sound effect. Each component can then be modified individually. When the helicopter is further away, only the sound of the rotor blades is heard, but as the helicopter approaches the sound of the engine and the tail rotor can be heard. The same technique could be applied to any moving vehicle. For a car, the sounds of wheels and chassis for example could be extracted when the car is further away, only playing the engine sound or only a noise approximation of the engine sound.

Because game worlds, in general, are growing in size, sound designers have to generate sufficient content to populate the worlds. This is a problem because samples must be stored in secondary storage, and when used, be brought from the secondary storage to the RAM. Therefore, it competes with graphics data computational resources. A sampled sound always uses the same amount of RAM as long as it’s playing, no matter how much it’s attenuated. A solution to this problem is some of the sounds can be replaced with procedurally generated, synthetic sounds. Research has shown how a synthetic sound source can fade in and out of a
sound scene, producing a variable load on the RAM. Procedural synthesized sound is also pure code, so it’s almost no data to store. (Farnell, 2007)

Synthetic sound is often criticized for not being too realistic, compared to its sampled counterpart and is therefore not supported with time and budget by developers. Farnell (2007) argues that procedural audio being 15 years behind in development. Extremely high sonic realism to match the graphics of the games of today is possible, but according to Farnell (2007), political obstacles, rather than technical, for further development must be removed.

1.2 Sound Synthesis Techniques
According to Moffat & Reiss (2018), the methods for synthesizing sound can be divided into three different categories: sample-based, signal-based, and physical models. The sample-based model involves audio recordings, cut and rearranged to produce new or similar sounds. The signal-based approach is based on the analysis and re-synthesis of the waveform of a real-world sound. Physical models are based on the physics of systems that generate sounds. These models represent vibratory patterns and acoustical properties. Moffat & Reiss (2018) further describe the very complex nature of complete physical models and argue that a model is considered better the more physics details are incorporated. Verron & Drettakis (2012), researching environmental sounds in video games, claims that a purely physical approach in video games is impossible. This is due to the heavy computational cost and the problems in creating a general model for a wide variety of environmental sounds. Böttcher & Serafin (2009) uses "physically inspired models" in their research, and defines it as "...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 (p. 1)." This might be a way around some of the problems like computational cost.

1.3 Physical Modeling
Physical modeling could be beneficial for the sound designer because of the variety of sonic alterations that could be obtained from using just one physical model. Hendry and Reiss (2010) researched the concept of replicating a large number of samples using a small number of physical models. The study aimed to find out if a sound effect could be synthesized with such accuracy so that it wouldn’t be distinguishable from the original recording. One of the physical models used in the study was based on a DC electric motor and controlled by a speed parameter. The model was originally created by Farnell (2010). To test out the hypothesis, two contrasting
samples were selected from a sample library, a household drill, and a small boat engine. The physical model of the DC engine was adapted to replicate the two samples. For increased realism, a way of humanizing the physical model with variations in the sound was introduced. A random walk was applied to the maximum speed, the volume of the brush sound, and the gain parameters for three bands in a parametric equalizer. The results show that spectral content between the samples and the physical models matched very closely. The zero-crossing rate however was not accurately represented among the physical models. In conclusion, the experiment was considered successful, and the authors suggest the methodology to be carried forward for other sound effects such as combustion motors. The authors also suggest combining this model with methods of feature extraction. As a starting point for the sound designer, a computer could extract parameter values for a sample such as relative harmonic content and fundamental frequency, to be used in the DC motor model, for transposing a sample to a synthesis sound. Hendry and Reiss (2010) discuss possible improvements for increasing the realism of the model and mentions the random walk applied to the physical model as a contributing factor to the perceived realness of the model. The authors refer to the study done by Böttcher & Serafin (2009) where they suggest granular synthesis as another method for further randomizing the sound.

1.4 Granular Synthesis
Granular synthesis is one technique that might achieve realistic results and also leave space for creative design. Granular sounds are short (around 1-100 ms) sound events (grains) that usually are distributed randomly over time and attenuated randomly to increase the variation of a sound (Siddiq, 2017). Paul (2011) suggests granular synthesis as a method to augment and improve sound for dialogue, sound effects, and car engines. Real-time synthesis is used in Grand Theft Auto V. A granular engine is used to synthesize the player's car. Additional oscillators, synced with the granular clock to correlate with the fundamental pitch of the car engine. The oscillators were used to add for example sub-harmonic frequencies to the granular engine (MacGregor, 2014). Nathan Blias describes in an interview by Walden (2018) the use of granular synthesis in the sound design of the title “The Crew 2”, and mentions that everything in the game had to be constantly dynamic. Granular synthesis was a helpful tool used to cut their engine recordings into a multitude of loops which could then be recomposed.

1.4.1 Large Scale Granular Synthesis
Vachon (2009) suggests large-scale granular synthesis as a method for avoiding tedium. Large-scale granular synthesis is the process of deconstructing more complex sounds into individual
layers so that variation can be attained. An explosion, for example, can be deconstructed into several parts; the explosive transient, the high-end crackle, the body, and the deep bass rumble. By mixing and matching different parts of different samples, randomizing pitch, volume, and attack, a large number of unique explosions can be created using only a few samples.

1.4.2 Concatenative Synthesis
Concatenative synthesis is related to granulation in the way it splits the input into short grains. But it also resynthesizes an audio input stream by categorizing, manipulating, and choosing the best grains for replicating the audio input (Paul, 2011). Schwarz et al. (2016) are perceptually evaluating three different concatenative synthesis algorithms for a wide variety of sounds including transportation sounds, mechanical gears, textures generated by human and natural elements (air, water, fire). The study demonstrates that the three algorithms in close performance in terms of quality and similarity to the original recordings.

1.5 Perceptual Evaluation of Synthesis Techniques
A lot of work has been done around sound synthesis evaluation, although according to Moffat & Reiss (2018) many evaluations have not considered realism or related attributes. The authors identified which synthesis methods could produce believable audio samples that could replace recorded sound samples, and used realism as the attribute for rating the synthesis methods. Schwarz (2011) reviewed 94 published articles on methods of sound texture synthesis and found out only 7 contained any perceptual evaluated results. Moffat & Reiss (2018) states that in cases when evaluations are carried through, they are rarely subjective, and even more uncommonly comparative, as when the proposed method is compared to alternative methods. There is according to Moffat & Reiss (2018) a lack of evaluation methodologies as well as a lack of an understanding of the current state-of-the-art in sound synthesis. In particular, that a single synthesis method is never expected to work in all cases, however, several synthesis methods might work for each case.

1.6 Auditory feedback from driving
Given the more intricate nature of driving, compared to other small engine sounds such as a household drill or small boat engine, the importance of auditory feedback from maneuvering a car must be considered. Baldan et al. (2015) discuss the perceptual and cognitive aspects of car driving and describes it as due to the cross-modal feedback a complex, yet natural activity where audio plays an important role. Referring to over a century of investigation in this area, the driver is exploiting visual, haptic, proprioceptive, and auditory cues to successfully
maneuver the vehicle and to effectively perceive its motion and behaviors. According to the authors, previous research has shown that sound contributes to shaping effective and convincing driving simulations. When the noise from the engine rotations provides the driver with a speed reference, making looking at the speedometer unnecessary. Because of this phenomenon, the usage of procedural vehicle synthesis has been given a relatively new area of applicability, in silent cars. Active Sound Design (ASD) is being used for both the exterior and interior in electrical vehicles, alerting pedestrians and giving the driver auditory feedback about the driving. If the political obstacles from developers previously discussed by Farnell (2007) are hindering the prospects of synthetic audio, the game sound designer might look at the potential in the rapid development of ASD for electronic cars for use in video games.

2. Research Question, Aim, and Purpose

Car engine sounds are complex sounds and can be difficult to recreate in video games. The traditional method of pitch shifting and crossfading sampled audio files certainly has drawbacks, not at least that it’s a time-consuming process. Granular synthesis and physical modeling are other alternative methods. Suggested physical models generally aim to incorporate as accurate physical representations as possible, but are in other research often referred to as physically inspired or physically-based models, since complex sounds are difficult to replicate the physics of in detail. Sample-based methods like granulation and concatenate synthesis are other suggested methods, producing realistic engine sounds. Which one produces the most realistic result is difficult to conclude since there are many variables to consider when designing and implementing, changing one variable in one model, for example changing the amplitude of an overtone or changing the way the pitch might be randomized, could generate a very different result for the listener. There is far from a standardized physical model for designing car engine sounds, but a lot of research is based on the work of Farnell (2010).

This work aims to evaluate three methods for creating realistic car engine sounds in the context of video games. By evaluating sound design methods relevant for its purpose, the result would provide a guideline for what method a sound designer could utilize.

To evaluate the realism of a proposed method for designing a sound, the definition of realism must first be considered. For this study, realism is rated based on two categories; interactivity (how the sound responds to player input) and sonic realism (the sonic attributes of the sound). These categories are going to be labeled in the study as responsiveness and perceived realism.
The labeled scale for rating realism in the main study are taken from the study on perceptual evaluation of synthesized sound effects (Moffat & Reiss, 2018).

2.1 Expected results
Synthesis methods for replicating real-world sounds are in general critiqued for sounding artificial, not yielding the same results as recorded audio. For car engine sound design, sample-based models and granular synthesis would be the most common methods used in the game industry. These two methods would be expected to be the most preferred since they are widely used in the industry. A physical model on the other hand would not be expected to be equally preferred, due to the method not being widely used in the game industry for car engine sound design.

3. Method
To evaluate the methods for creating realistic car engine sounds, three models were constructed, each one representing a sound design method. A sample-based model was compared to a granular synthesis model in a driving simulator game, followed by a post-study where a physical model was compared to the previous models. After completing the driving simulator, the subjects were asked to complete a survey in which the two models in the driving simulator were rated on a labeled scale based on perceived realism and a numbered scale based on responsiveness. Subjects were also asked to describe if the driving experience felt better in one of the cars, and why so.

3.1 Demographics
Since the sounds were easily identifiable from each other, experienced listeners such as sound engineers or musicians weren't a necessary subject pool for evaluating the sounds. Instead, the subject pool consisted of "people who play video games", with no further defined requirement needed to participate. The reason was that subjects had to have some kind of reference to video game sound design, preferably vehicle sound design in video games, to make a valid assessment. Subjects were mostly university students in various fields. No other demographical data were collected other than about prior video gaming experience.
3.2 Pre-Study
To gain feedback on the experiment and ensure the driving simulator worked properly, a pre-study was done. Participants were university students with prior gaming experience, some of them with experience in game development. The Pre-Study included the driving simulator, the questionnaire, and the post-study listening test. The pre-study was, like the main study followed
through over distance. Small adjustments were made to the driving simulator, such as removing development messages that occasionally appeared on the screen.

3.3 Experiment design
A simple driving simulator was created in Unreal Engine (Epic Games, 2020), based on Unreal’s "vehicle template” map. The map was slightly enlarged to provide enough space for the vehicles to reach maximum speed. Two cars based on Unreal’s "vehicle template car” blueprint were used, but with the physical mesh of the sports car in Unreal's Vehicle Variety Pack (Switchboard Studios, 2018). The physical mesh of the sports car was a basic, not very detailed visual representation, chosen to represent a generic sports car and not a specific car model. Thereby avoiding potential preconceived notions, the informant could’ve had about the car’s sonic attributes when visually appearing on screen.

3.3.1 Turbomania
The driving simulator was given the name ”Turbomania”. After opening the driving simulator, the subject was introduced to the main menu with a start button, a quit button, a volume slider, and a music track playing. The volume slider in the menu was intended for the player to adjust the volume of the music before entering the driving simulation. The player would start the game as a third-person character facing two versions of the same car, a red and a blue car. The player could choose which car to drive by walking up to it and enter the car. The player could at any time exit the vehicle and enter the other one. The driving simulator had no time limit, but the player would decide when to quit.

The car had five gears, which were automatically transitioned in between when accelerating or decelerating the car. No visual representation of the current gear or RPM was shown so that the subject’s only indication of those parameters came from the auditory feedback. Figure 5 shows the float values from the car movement being converted into RTPC float values corresponding to kilometers/hour, which was used to set the RPM-value in Wwise.
Figure 3: Shows the start position once starting the driving simulator

Figure 4: Shows the in-car view
Figure 5: Blueprint showing the car’s speed controlling the engine sound RPM-value

Figure 6: Blueprint showing the logic behind entering and exiting the car

Figure 6 shows the logic behind entering and exiting the car. When the player was standing next to a car, he could choose to possess the car. This action would take him to the view shown in figure 4 and immediately trigger the engine idle to start. The third-person character would then disappear on the screen. When choosing to exit the car, the third person character would then appear on the screen right next to the car’s current position, being possessed by the player controller as shown in figure 3. This action would also stop the engine sound.

3.4 Choice of Method

Sound designer Stephen Baysted is interviewed in Donnolly (2014), talking about the titles "World of Speed" and "Project Cars". For those titles, well over a hundred separate wave files were used for each car, cross-faded and layered to be ‘played’ by physics input. This is a real-world example of the great magnitude of audio files a racing game could utilize. For this study,
a simplified model was incorporated, although it still had to be a valid representation of the method that is sample-based car engine sound design.

To construct an experiment where two sound design methods could be assessed and be compared against each other, the methods used were suggested by Audiokinetic (2021) as two appropriate methods for reproducing car engine sounds in Wwise (Audiokinetic, 2020). Those being a sample-based method using the blend container function in Wwise (Audiokinetic, 2020), and a granular synthesis method using the plugin REV (Crankcase Audio, 2021).

3.4.1 Blend Container

The blend container function was used to pitch-shift and crossfade recorded audio files of a Porsche 911. A total of 13 looped assets were imported, each asset designed to represent a static RPM value. Six assets were used for the engine revving up and six for the engine revving down, one asset was used for the engine idle. Each asset was assigned a pitch curve attached to an RPM game parameter. This pitch curve was based on a "build smart pitch curve"-function, and manually modified to correspond with the shifting of gears. The assets were crossfaded into each other for smooth transitions when the RPM game parameter changed.

3.4.2 REV

REV (Crankcase Audio, 2021) works as a plugin for game engines and game audio middleware, solely for designing vehicle engine sounds. It works in the frequency domain to allow variable speed playback in real-time of recorded accelerations. It tracks the harmonic components of an engine accelerating or decelerating and granulates the recording into discrete slices of individual pistons firing. The "ramp" can then be reproduced as it was originally recorded. But it can also be reproduced at a slower or faster rate, by skipping over cycles or duplicating and repeating cycles.

According to Donnolly (2014), interviewing the creators of REV (Crankcase Audio, 2021), the game industry was in urgent need of a better vehicle engine modeling system. The main goal of the system was to avoid the artificial sound that usually occurs when using a traditional loop-based method of pitch-shifting recordings. Another goal according to the creators was reducing the complexity and time required to create a high-quality-sounding engine model.

REV (Crankcase Audio, 2021) was used as a plugin in Wwise to represent a granular model. A preset of a Porsche 997 was used, with the RPM parameter attached to the RPM game parameter.
3.5 Additional Sounds

For ecological validity to the experiment, several additional sounds were implemented in the driving simulator. Figure 7 shows the logic behind the drift/handbrake sounds. The variety of sound assets would trigger if the player used the handbrake over a certain speed, or if an offset value corresponding to turning left/right was exceeded while the car was exceeding a certain speed.

- Ambiance (light gusts of wind to provide a lifelike environment)
- Footsteps (for the third person character)
- Drift/Handbrake (when exceeding a certain speed, a variety of slip sounds are played when using the handbrake or when tires are drifting)
- Wall impact (when the car touches the wall over a certain speed, versions of impact sounds are played, depending on the speed of the car)
- Menu music (to adjust the main volume)
- Slider UI (when using the volume slider in the pause menu to provide the player with auditory information about the volume)

Figure 7: Blueprint showing the logic behind the drifting/handbrake sound implementation

3.6 Tools

- Unreal Engine 4 (Epic Games, 2020) – Game Engine
- Wwise (Audiokinetic, 2020) - Game Audio Middleware
• REV (Crankcase Audio, 2021) - Granular Vehicle Sound Design Plugin
• Pure Data (Miller Puckette, 2020) - Visual Scripting Programming Language
• Reaper (Cockoc Inc. 2021) – Digital Audio Workstation
• EaseUS RecExperts (EaseUS, 2021) - Recording Software
• Keyboard and Mouse
• PC

3.7 Porsche 997 and 991
The two models in the driving simulator were based on recordings of a Porsche 997 and a Porsche 991. The 997 is part of the 911 series and both cars have a flat-6 engine. The description of the 997 model from Crankcase Audio (2021) states: "Porsche 997 a.k.a the 911, this has the distinctive flat-6 sound of the preeminent German sports car". Based on this, and the high quality of both recordings, they were considered to be enough alike to be used for the granular synthesis model and the sample-based model.

3.8 Main Study
16 subjects participated in the main study. Participating subjects received a .zip-file containing the driving simulator and a text document with instructions which can be seen in Appendix C. After completing the driving simulation, they were asked to fill out a questionnaire where the post-study listening test was included. The study was completely followed through over distance, meaning the subjects used their PC, keyboard, mouse, and listening equipment. Since the assessment was based on rather great sonic differences and assessment of responsiveness from player input, the subject’s native environment was preferred. The ecological validity here was assumed a better condition to preserve, rather than equal listening conditions. The assumption is that different playback systems would not have much of an impact on the result.

3.8.1 Main Study Survey Questions
After completing the driving simulator, all subjects were asked to complete a survey on google forms with the post-study listening test included in the survey. The survey included the questions down below in the following order:

How often do you play video games?

• Every day
• A few times a week
• Once a week
Do you play video games containing driving elements?

- Often
- Sometimes
- Occasionally
- Rarely/Never
- I don’t know

How realistic did you perceive the sound of the blue car?

- Very Realistic
- Quite Realistic
- Quite Unrealistic
- Very Unrealistic

How realistic did you perceive the sound of the red car?

- Very Realistic
- Quite Realistic
- Quite Unrealistic
- Very Unrealistic

How responsive to your steering and input did the blue car’s engine sound feel?

The subjects were here asked to rate the responsiveness on a scale 1-5.

How responsive to your steering and input did the blue car’s engine sound feel?

The subjects were here asked to rate the responsiveness on a scale 1-5.

Overall, did the driving experience feel better in one of the cars? Then why so? (Free text area)

3.8.2 Post-Study Listening Test Survey Questions

Instructions: You will now listen to three sports car sounds. Click on the link to listen to the sounds: (link to a google drive with the three audio files).
In a driving simulator like Turbomania, which sound would you prefer for the car? Rank the sounds in your preferred order with the most preferred first, followed by your second and third choices. If two or more are equally preferred, then write "_and_are equally preferred" or "all are equally preferable". (Free text area)

If any one of the sounds (X, Z, Y) would offer a better driving experience than the other two, then why is your preferred sound preferable to the other two? (Free text area)

3.9 Post-Study Listening Test
The post-study listening test was constructed to include an evaluation of the physical model and compare it to the granular synthesis model and the sample-based model. The physical model was initially set to be included in the driving simulator, but due to technical issues implementing the patch in Wwise, the post-study listening test was constructed.

The physical model was a Pure Data patch shown in figure 8 and was constructed by Farnell (2010). The subjects were to listen to three short audio recordings of the sample-based model, the granular synthesis model, and the physical model. They were then asked to rank the audio files, based on preference in a described context of the audio files being used for a sports car in a game like Turbomania. They were also asked to motivate why their preferred sound would offer a better driving experience than the other two. The audio files in the post-study listening test were all set to -20.6 LUFS and recorded in Wwise (Audiokinetic, 2020) by manually adjusting the RPM parameter from minimum to maximum value and vice versa. The Pure Data patch was recorded in the same way but in Pure Data.

A total of 14 people participated in the post-study listening test. Two subjects misinterpreted the post-study listening test survey and ranked the car models in the main study instead, these answers weren’t included in table 3 shown in the results section. Some subjects ranked two sounds as equally preferred, those models have both been given the same number in ranking.

Since the post-study listening test was designed primarily to evaluate the Physical Model in comparison to the other models, conclusions about the sample-based and the granular synthesis model in comparison to each other will function mainly to support the results from the main study.
Figure 8: The Pure Data patch used for the physical model
4. Results and Analysis

To summarize, results from the main study showed that the granular synthesis model was considered the most realistic. Ratings of responsiveness showed no statistical significance. Data from the post-study listening test showed a significant preference for the granular synthesis model as the most preferred and the physical model as least preferred. The perceived realism ratings were checked against the demographics data to see if something in the subject’s experience potentially had an influence, specifically, the mean values for every demographic pool rating were checked. However, no correlations were found.

The quantitative data are presented in the tables down below. Responses on perceived realism, responsiveness, and analysis of the demographic data were analyzed for statistical significance using the t-test calculator (GraphPad Software, 2021). For the analysis of the qualitative data, the subject's responses have been coded to “blue car” and “red car”, and as either positive or negative attributes. These responses are in table 1. Answers in Swedish have been translated to English.

4.1 Main Study Survey Responses

![Survey responses on perceived realism for the granular synthesis model](image)

Figure 9: Survey responses on perceived realism for the granular synthesis model
Figure 10: Survey responses on perceived realism for the sample-based model

The results of perceived realism ratings on the Granular Synthesis Model ($M = 2.44$, $SD = 0.63$) and the Sample Based Model ($M = 2.19$, $SD = 0.83$) indicate that the subjects perceived the Granular Synthesis Model as more realistic, $t(15) = 2.2361$, $p = .0410$.

The comparisons of perceived realism were given values between 1 (= Very Unrealistic) and 4 (= Very Realistic) for data analysis.

Figure 11: Survey responses on responsiveness for the granular synthesis model
There was no significant preference on responsiveness, \( t(15) = 1.1593, p = .2644 \), despite the Granular Synthesis Model (\( M = 3.19, SD = 0.91 \)) attained a higher score than the Sample Based Model (\( M = 2.88, SD = 1.15 \)).

**Table 1: Attributes categorized as positive and negative from the survey question: “Overall, did the driving experience feel better in one of the cars? Then why so?”**

<table>
<thead>
<tr>
<th>Blue Car (Granular) - Positive attributes</th>
<th>Blue Car (Granular) - Negative attributes</th>
<th>Red Car (Sample-Based) - Positive attributes</th>
<th>Red Car (Sample-Based) - Negative attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credible gear changes</td>
<td>Awkward shifting of gears</td>
<td>Organic</td>
<td>Jumping inbetween sets of sounds</td>
</tr>
<tr>
<td>Responsive engine sound</td>
<td>Sluggish</td>
<td>Realistic response</td>
<td>Thin in lower speeds</td>
</tr>
<tr>
<td>Depth to the sound in top speed</td>
<td>Muffled</td>
<td>Better indication on steering</td>
<td>Tiresome after a couple of minutes driving</td>
</tr>
<tr>
<td>Realistic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smooth overlays</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impactful sound</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fits the car model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better on higher speeds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better when turning</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1 shows the attributes being applied to both car models. The attributes do not show any trends for each column. However, a few trends can be seen throughout all columns as some attributes are more frequently apparent. Attributes describing steering and responsiveness can be considered related to each other, such attributes are together appearing four times. Attributes describing the sounds in terms of speed are as well appearing four times. Attributes on gear changes are appearing three times. Based on this, these aspects could be considered important for creating realistic car engine sounds, as they seem to make the driving experience feel better or worse. The full answers can be found in Appendix A.

4.2 Main Study Analysis

A conclusion can be drawn from the main study about the perceived realism of the granular synthesis model being superior, although not regarding responsiveness. The reasons for these perceptions are unclear because the qualitative data presented in table 1 shows a great variety of attributes being applied on both models, making further analysis difficult. However, a few attributes from the survey showed the techniques behind the two methods, such as the granular synthesis sounding smooth in its gear transitions. This could be seen as an advantage of using the technique.

The sample-based model would be expected to sound more realistic with fewer artifacts. Attributes like “organic” and “less broken” are not unexpected descriptions for this model, as well as “jumping in-between sets of sounds”. One subject described the sample-based model as “tiresome after a couple of minutes”, which could also be an expected downside of using sample-based audio in video games.
4.3 Post-Study Survey Responses

Table 2: The multitude of each ranking for each model in the post-study listening test. The models were ranked as first, second and third according to preference. The table shows the ratings for each model in ascending order.

<table>
<thead>
<tr>
<th>X (Sample-Based Model)</th>
<th>Z (Granular Model)</th>
<th>Y (Physical Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Mean Value: 1.857142857  1.153846154  2.692307692
Median: 2  1  3
Standard Deviation: 0.770328887  0.375533808  0.630425172
Variance: 0.593406593  0.141025641  0.397435897

A chi-square test of goodness-of-fit was performed to determine whether the number one rankings were equally distributed. Number one rankings for the three sound design models were not equally distributed in the population, $X^2 (2, N = 17) = 8.941, p < .05$. Meaning it does appear to be a statistically significant preference.

Another chi-square test of goodness-of-fit was performed to determine whether the number three rankings were equally distributed. Number three rankings for the three sound design models were not equally distributed in the population, $X^2 (2, N = 13) = 12.154, p < .05$. Meaning it does appear to be a statistically significant preference.

The chi-squared tests were calculated using the Chi-Squared Calculator for Goodness of Fit (Social Science Statistics, 2021).
Table 3: Attributes from the post-study listening test survey about why one sound would offer a better driving experience than the other two according to subject’s preference, categorized into positive and negative attributes.

<table>
<thead>
<tr>
<th>Z (Granular Synthesis) - Positive Attributes</th>
<th>Z (Granular Synthesis) - Negative Attributes</th>
<th>X (Sample Based) - Positive Attributes</th>
<th>X (Sample Based) - Negative Attributes</th>
<th>Y (Physical Model) - Positive Attributes</th>
<th>Y (Physical Model) - Negative Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gives a sense of engine power</td>
<td>Realistic</td>
<td>Fake</td>
<td>Better quality</td>
<td>Fake</td>
<td></td>
</tr>
<tr>
<td>Changes according to engine revolution</td>
<td>Not too bumpy</td>
<td>Fake</td>
<td>More seamless</td>
<td>Artificial in the way its pitch rose</td>
<td></td>
</tr>
<tr>
<td>Realistic</td>
<td>Clear progression</td>
<td>Looping on a granular level</td>
<td>Less broken</td>
<td>Fake</td>
<td></td>
</tr>
<tr>
<td>Not too bumpy</td>
<td>Authentic</td>
<td>Barely any structure to it</td>
<td>Looped on a granular level</td>
<td>Synthetic</td>
<td></td>
</tr>
<tr>
<td>Like a real car</td>
<td>Less “jumpy” acceleration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growling</td>
<td>Better quality</td>
<td></td>
<td>Low-passed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Realistic</td>
<td>More seamless</td>
<td></td>
<td>Old game-sounding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural acceleration changes</td>
<td>Less broken</td>
<td></td>
<td>Parody of sports car-sounding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More nuance</td>
<td></td>
<td></td>
<td>No high frequencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rattling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detailed engine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mix of both highs and lows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More energy in the lower register</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 shows the attributes from the post-study listening test, categorized into positive and negative attributes for each model. The results here show that the greatest number of negative attributes were applied to the physical model and the greatest number of positive attributes were applied to the granular synthesis model. The full answers can be found in Appendix B.

4.4 Post-Study Analysis

The granular synthesis model was considered the most preferred in the post-study listening test with statistical significance. The physical model was the least preferred. An observation from the data shown in table 3 is that no negative attributes were applied to the granular synthesis model, and the granular synthesis model was given the largest number of positive attributes. Negative attributes applied to the physical model indicate that the limitations of the physical model were noticeable to the subjects. However, some attributes regarding possible advantages were as well applied. Since the physical model is not based on manipulating an audio recording in any way, contrary to the two other models, the attributes describing the quality of the sound as positive could indicate potential advantages with the method.

4.5 Demographic data analysis

![Means of Realism ratings divided into frequency of playing video games](image)

*Figure 13: Means of subject’s ratings of realism for each model divided into demographic pools*
Multiple t-tests were conducted using the t-test calculator (GraphPad, 2021). They showed no significant difference between ratings in any of the demographic pools, nor in ratings of the same model between demographic pools, despite the slight tendency of subjects who rarely/never played video games containing driving elements rating both models higher. The small number of participants in the study makes concluding the influence of demographic traits difficult.

5. Discussion

As seen in the results section above, the granular synthesis model was rated the most realistic model in the driving simulator. It was also the most preferred in the post-study listening test. This is an interesting result because using granular synthesis for this purpose could offer some potential advantages for the sound designer. Firstly, it can be a less time-consuming process for designing vehicle engine sounds, in contrast to the sample-based method. Secondly, making use of only one audio file for the complete design of one car engine, as in this experiment, could potentially reduce the amount of RAM being used. This as opposed to implementing 13 audio files for this experiment, or well over a hundred as stated in the previous example in the method section (Donnolly, 2014).

Since the quantitative data collected about responsiveness in the main study showed no statistical significance, few findings can be discussed here. One possible explanation for these results could be that the responsiveness of these two implementations was fairly equal, did not reveal potential differences in these models or the methods, and therefore did not have much
of an impact on the driving experience. However, this is unlikely because the qualitative data showed positive and negative attributes regarding responsiveness, for example, re-occurring attributes describing the responsiveness of the two models, which indicate that the responsiveness of the sound was a contributing factor to why the driving experience felt better in one of the cars. Another possible explanation could be confusion about what attributes to assess when evaluating responsiveness. The term could potentially be misleading to subjects and subjective interpretations could have led to different subjects evaluating different aspects of the term. Realism in contrast could be seen as a term for which a greater consensus prevails. A description of responsiveness could have been included in the survey to avoid this potential error source.

5.1 Critique of the Method
A few things could have been made different in the design of the method. An ideal scenario would have been to use the same recording for both the sample-based and the granular synthesis model, to minimize the differences between them. However, the physical model isn’t based on a specific car model but functions more like a template that a sound designer could utilize for motor engine sound design. Nevertheless, it might have yielded different results if, for example, the granular synthesis would have used the same recordings as the sample-based model. Furthermore, since the main study was partly set up to evaluate the perceived realism of the two sound design techniques, a more appropriate attribute might have been to evaluate their believability. Because realism isn’t always preferred in video game sound design, but rather a believable sound in combination with the visuals.

Another thing is that the car sound design in the driving simulator could have been more detailed for both cars. For example, chassis sounds and sounds from the car rattling weren’t implemented, which could’ve made the driving experience in both cars more realistic. A more detailed implementation of additional sounds for both cars could have made both cars being perceived as more realistic. However, a very detailed implementation of additional sounds could have drawn the subject’s attention away from evaluating the engine sound.

5.2 Pure Data
Physical modeling as a choice for the implementation of vehicle engine sounds would still be considered a rarity in the game sound design industry at this time. Since engine sounds are complex sounds to physically model, there is a lack of synthesized models available for the sound designer. The initial plan for the experiment in this study was to include the Pure Data
(Miller Puckette, 2020) patch into the driving simulation, to test it in an in-game scenario. By using the Heavy Compiler (Github, 2018) to compile the patch into C source code, a plugin in Wwise (Audiokinetic, 2020) could then have been constructed. Other patches were successfully implemented using this method but failed with the patch intended for this experiment. After debugging and trying to optimize the patch for the compilation, the patch still wasn’t compiled successfully, with major artifacts playing it in Wwise (Audiokinetic, 2020). The pure data patch was instead included in the post-study listening test but was assessed without any visuals. A possibility is that the results could have been different if the patch was evaluated in the context of appropriate visuals.

5.2.1 The Heavy Compiler
The Heavy Compiler (Github, 2018) is an open-source audio programming language compiler that generates C/C++ code. It was created to leverage Pure Data as a design interface for creating interactive music and sound products. The Heavy Compiler (Github, 2018) used to be a supported program by the developers but is currently without a maintainer and only exists as open-source code. This is relevant here because the heavy compiler is one of the very few examples of software available to compile visual scripting sound design patches like patches in Pd into plugins that can be implemented to game engines or game audio middleware. For the average sound designer, this makes the feasibility of using, for example, Pd to create physical models low. Of course, a large game studio with game programmers may not have the same issues authoring a plugin. However, most companies will not allocate those programming resources to audio.

Pd patches have been widely used in other research on physical modeling of less complex sounds than car engines, like environmental sounds and foley, with successful results. If sound designers would use physical modeling for video games, there needs to be reliable software that can compile visual scripting patches into plugins. Or sound designers would need a lot of programming skills and knowledge about physical modeling to further develop the use of Pd for car engine sound design today.

5.3 Future Research
Even though the results of this study didn’t indicate any preference towards the use of a physical model for car engine sound design, the physical modeling method can certainly be improved, and it would be interesting to research other physical models besides engines. As stated in the introduction section, Farnell (2007) argues that physical modeling can be used for
distant sound objects, not as central in the mix as the cars in this experiment were designed to be. It would be interesting to research such a scenario regarding car engine sound design in video games. Furthermore, measurements on the amount of RAM being used for all techniques used in this study would also be an interesting parameter to include in future research. Since a granular synthesis model could also potentially make use of less RAM than a sample-based model, this is also an area where future research could do measurements.

5.4 Conclusion

This study has evaluated the driving experience of three sound design methods for creating car engine sounds in video games. The results have shown the method using granular synthesis to be perceived as the most realistic, however not the most responsive. The method using physical modeling has been the least preferred.
6. Bibliography


https://github.com/enzienaudio/hvcc


Appendix A: Full answers regarding if the driving experience felt better in one of the cars and why so.

<table>
<thead>
<tr>
<th>Blue Car – Granular Synthesis</th>
<th>Red Car – Sample Based</th>
<th>None/Equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>The blue car felt a little better overall. The gear changes sounded a lot more credible, and the engine sound felt more responsive.</td>
<td>In the red car. It felt more organic sounding.</td>
<td>I did not really feel a difference, but I almost never play driving games</td>
</tr>
<tr>
<td>I think I preferred the blue car because of it having more depth in the sound while in top speed.</td>
<td>Red car felt better. The Blue car felt a bit sluggish.</td>
<td>No, they both sounded good</td>
</tr>
<tr>
<td>The blue car was more realistic. Despite that I think none of the cars was especially realistic when shifting gears. It felt like the car was going top speed on each gear, i.e. high rpm-values are reached before the gear shift happens.</td>
<td>The response in the red car felt more realistic.</td>
<td>Blue felt better on higher speeds and when turning, red felt better on lower speeds. The blue car's suspension felt better.</td>
</tr>
<tr>
<td>The blue car felt better than the red car. The red car felt as if it was jumping inbetween set sounds, and the blue car felt more &quot;smooth&quot; with it's overlays. At the same time though, the acceleration of the blue car felt awkward as the car was</td>
<td>The red one I think. Had to think hard before making up my mind, but the red one sounded different when steering, whereas the blue one sounded more like simply muffled (I think, hard to tell). Also, I managed to tumble the</td>
<td></td>
</tr>
</tbody>
</table>

"shifting gears" - so none of them felt realistic, but the red car felt "sampled" to the inputs I was making.

<table>
<thead>
<tr>
<th>damned blue one quite a lot while with the red one, the sound gave me better indications on how I handled the car while driving and therefore made it easier to sense when I was at risk of losing control. Further note: Why I felt both to be unrealistic, was due to that the engine sounds a lot different on real lamborghini (these reminded me of older games), but the main reason was how the RPM's sounded in each gear in relation to speed. I hit TAB and saw in 1st person view both the speedometer and gear and found the sound to be inconsistent with what I saw there.</th>
</tr>
</thead>
</table>

Enjoyed the blue car more. More impactful sound that I think fits the car model better. The red sounded less broken which fit the look of the cars better.

**Appendix B: Full answers from the post-study listening test survey regarding if any of the sounds would offer a better driving experience according to the subject’s preference**
Why any of the sounds (Z = Granular, X = Sample-Based or Y = Physical Model) would offer a better driving experience according to the subject's preference

<table>
<thead>
<tr>
<th>Sound</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>I prefer the Z sound because it gives a lot more sense of engine power. And I think it also changes the most depending on how you revv the engine.</td>
</tr>
<tr>
<td></td>
<td>The red car sounded a little bit more thin in the lower speeds, I felt like I got tired of that sound after driving around for a couple of minutes.</td>
</tr>
<tr>
<td>X</td>
<td>X and Z feels more realistic, not to bumpy.</td>
</tr>
<tr>
<td></td>
<td>X felt like it had a clear progression from first to last gear. Only by listening to the sound you could feel the car going faster and faster, and I think that's the key point of designing sound for games, to make the sound emphasize the feeling the gameplay tries to create, rather than to sound exactly like something in real life.</td>
</tr>
<tr>
<td></td>
<td>It sounds more like a real car, it sounds almost like it is sampled from a real car. The other two just sounds fake.</td>
</tr>
<tr>
<td></td>
<td>Y sounded artificial in the way that its pitch rose. I found Z and X to be equally preferable but they just sounded like different car models. They had qualities which made them sound real and there was nothing that really struck me as sounding unnatural.</td>
</tr>
<tr>
<td>Y</td>
<td>Z sounded best, it has more of the growling sound I associate with cars.</td>
</tr>
<tr>
<td></td>
<td>Z sounded most realistic. The changes between the different acceleration stages felt more natural. X and Y sounded fake and like they were looping on a granular level. Y was worst for this.</td>
</tr>
<tr>
<td></td>
<td>Sound X sounded most authentic.</td>
</tr>
<tr>
<td></td>
<td>Y definitely sounds the worst. Y just sounds synthetic and low passed - it doesn't even really sound like a car.</td>
</tr>
<tr>
<td></td>
<td>Between X and Z it is pretty equal to me. I don't have a strong preference towards any of them, but X has a less &quot;jumpy&quot; acceleration, while Z has more nuance to the actual sound of the car, such as the car ratteling and a more detailed engine.</td>
</tr>
<tr>
<td></td>
<td>Z sounded slightly more realistic. X sounded closest to what you were using (not bad, but unrealistic). Y sounded like a very old game or a parody of a sports car.</td>
</tr>
<tr>
<td>Z</td>
<td>z sounds more natural than x and y.</td>
</tr>
<tr>
<td></td>
<td>Z seems to have a mix of both high and lows, which X and Y had mostly one over the other. Z had more energy in the lower register, but it still had some high frequencies as well. Where as Y had no high frequencies at all, and X had barely any structure to it.</td>
</tr>
</tbody>
</table>
They sounded more seamless, thus less broken. Y and Z had better quality in my opinion but both had downsides.

Appendix C: Full experiment instructions for subjects

In this experiment, you're going to play the driving simulator "Turbomania. After you've played, you will be asked about how the sound of each car contributed to the driving experience. You can always quit the experiment at any time.

Instructions:

Press the ENTER-key when standing close to a car to enter it. Press ENTER again to exit the car.

Handbrake using the Space Bar

Navigate using WSAD-keys or the arrow-keys

ESC-key to enter pause menu

Adjust the volume using the slider in the main menu or pause menu

Start over by clicking "Start Over" in the pause menu. You can always start over when you feel so.

Step 1: Open the Turbomania.exe file

Step 2: Play the driving simulator for as long as you like

Step 2: Answer the questionnaire: (link to questionnaire)