

Perceived audio quality of compressed audio in game dialogue

Anton Ahlberg

**Ljudteknik, kandidat
2016**

Luleå tekniska universitet
Institutionen för konst, kommunikation och lärande

Perceived audio quality of compressed audio in game dialogue

Anton Ahlberg

Bachelor's thesis, Audio Engineering

*Luleå tekniska universitet,
Department of Arts, Communication and Education*

Abstract

A game could have thousands of sound assets, to fit all of those files to a manageable storage space it is often necessary to reduce the size of the files to a more manageable size so they have to be compressed. One type of sound that often takes up a lot of disc space (because there is so much of it) is dialogue. In the popular game engine Unreal Engine 4 (UE4) the audio is compressed to Ogg Vorbis and has as default the bit rate is set to 104 kbit/s. The goal of this paper is to see if untrained listeners find dialogue compressed in Ogg Vorbis 104 kbit/s good enough for dialogue or if they prefer higher bit rates. A game was made in UE4 that would act as a listening test. Dialogue audio was recorded with a male and a female voice-actor and was compressed in UE4 in six different bit rates. 24 untrained subjects were asked to play the game and identify the two out of six robots with the dialogue audio they thought sounded the best. The results show that the subjects prefer the higher bit rates that were tested. The results were analyzed with a chi-squared test which showed that the null-hypothesis can be rejected. Only 21% of the answers were towards UE4's default bit rate of 104 kbit/s or lower. The results suggest that the subjects prefer dialogue in higher bit rates and UE4 should raise the default bit rate.

Table of content

Page nr.

1. Introduction	4
1.1. Background	4
1.2. Game Industry	4
1.3. Dialogue in games	5
1.4. Speech intelligibility	5
1.5. Audio in games	6
1.6. Audio quality	6
1.7. Audio compression	6
1.8. Unreal Engine	7
1.9. Compression of audio in games	8
1.10. Purpose	8
2. Method	9
2.1. Listening test	9
2.2. Development and design	9
2.3. Stimuli	12
2.3.1. Dialogue	13
2.3.2. Compression	14
2.4. Listening conditions	15
2.5. Pilot study	15
2.6. Subjects	16
2.7. Questionnaire	16
3. Results	18
3.1. Comments from subjects	20
4. Analysis	21
4.1. Comments	21
4.2. Quantitative data	21
4.3. Chi-squared test	21
5. Discussion	23
5.1. Reflections on the results	23
5.2. Critique of the method	23
5.3. Subjects	23
5.4. Graphics and audio influence	24
5.5. Ecological validity	24
5.6. Expected results	24
5.7. Dialogue assessment	24
5.8. Optimizing compression	25

5.9.	Conclusion	25
5.10.	Further research	26
6.	Acknowledgements	27
7.	References	28
8.	Appendix 1	XXX
9.	Appendix 2	XXXIII

1. Introduction

Games are limited by physical media and data storage space that limits the amount of assets and sonic data we can have in a game. Available CPU limits how those sound assets may be processed and rendered. A game could have thousands of sound assets and to fit all of those files to a manageable storage space it is often necessary to reduce the size of the files to a more manageable size so they have to be compressed. One type of sound that often takes up a lot of disc space (because there is so much of it) is dialogue. Interestingly, recorded dialogue is very robust to compression. There are different types of compression in audio. The most commonly used are Mp3, Mp2, AAC, Ogg Vorbis and Opus. They all have varying impact on the sound. However, the price of storage has decreased and the computers are becoming faster with more processing power. Do games still need to compress audio and sacrifice quality? This study investigates the application of compression of in game dialogue. Because dialogue makes up a large amount of the audio in a game this test will investigate whether untrained listeners can discriminate high bit rates from lower bit rates and what bit rates they prefer.

1.1 Background

Collins (2013) says that *Halo 2* from Bungie had 16,000 lines of dialogue in the game and Chandler (2009) says that a game easily can have over 100 000 lines of dialogue. *Fallout 4* have over 111 000 lines of dialogue (IGN, 2015). It won't be long before we have games whose content surpass 200 000 lines of dialogue. Compressing audio does more than solve the problem of fitting the assets into the available storage space. For example, because assets require less space, more variations of the assets can be used. With different variations of sound assets and dialogue you can make the game come alive, feel less repetitive and prevent boredom when playing. This can be in form of multiple versions of the same assets or just new assets to create greater variety. With more storage the sound designer can get more creative when creating assets. Storage space may also impact sound quality if the sound designer can store more data the sound designer may be able to use higher bit rates and therefore may improve the sound quality. Dialogue is not only story based that only need one version, dialogue includes shouts, grunts and other human sounds that need variety. Computers get faster every year but we still compress audio in games as standard practice.

1.2 Game industry

For as long as we've been using computers and been playing games on them storage has become cheaper and the computer become faster. Today you can get 1TB of internal HDD storage for under 50 dollars or 45 €. There are no signs of that trend to stop, but why do games still being developed with low quality audio, compared to graphics. Sims (2014) says that many modern titles come with so pre rendered full motion videos (PRFMV) that can take up a lot if not most of the storage space, these are files with high quality video and audio. When our standards for video always rises from analog, to DVD (480p), to Blu-Ray and HD (1080p), to now UHD and 4k we have not seen the same development in audio. Even though there has been huge advancements in the last decades the audio branch of gaming are still limited by storage and CPU. Alten (2014) states that we will always be limited by CPU and storage but for how

long? For how long do we have to compress audio just to make it smaller? Boer (2002) says that compressed audio may have an impact on CPU usage, that compressed audio may use more CPU when decoding than PCM audio depending on cache allocation and the codec. Now that storage is cheaper couldn't it be a good trade to sacrifice some storage to make the CPU use more efficient. 52% of all game sales in 2015 were digital sales according to the Entertainment Software Association (ESA, n.d.). Less than half of all the game sales are limited by a physical disc when sold. That trend is probably continue as long as the supply in the online stores like Steam and EA Origins stays high. When looking at gaming through the years the demand has always increased (except for after the E.T. game for the Atari 2600, 1982) and will probably continue to do so. Entertainment Software Association has also released a study that says that the average gamer is 31 years old and plays games on a console (ESA, n.d.). Which means that the average gamer today is not a child that will grow out of his or her toys and gaming is here to stay.

A concern from many users that digitally download their game is that all languages are delivered. That means that there may be instead of 1 GB of audio it may become 5 or 10 GB. And if we then take into account that many of the aforementioned PRFMV come in many different languages this quickly becomes an issue. To only have to print one physical copy instead of one for each language is very reasonable because it is very expensive to print Blu-Rays and DVDs. But to not offer the 52% of all gamers that download their games the option to choose before they download which language they want it in seems unnecessary. They could even provide the option to choose multiple languages to download. If we compress audio to make more room for other content, why do the developers use so much storage for different languages? If a lack of storage and large file sizes is the problem and the reason we still compress audio it would be reasonable to see what takes up the most disc space. If we could find a way to get PRFMVs smaller and in one language the issue of space may become obsolete.

1.3 Dialogue in games

Dialogue in games, just as in movies, is very important for storytelling. Dialogue is a very easy way for a game to inform the player about what's going on and what needs to be done. If the quality of dialogue is compromised there is a big risk that the suspension of disbelief is sacrificed. Suspension of disbelief can be defined that the player can believe the unbelievable and sacrifice logic and realism for the sense of entertainment (Suspension of disbelief, n.d.). Immersion in games is when the player feel immersed and the player is captivated by the game. Not just surrounded but the player feel very invested in the story and the game world. If you as a player don't understand what is being said in a game, the story might not be communicated and the game's overall quality might be compromised.

1.4 Speech intelligibility

Darwin (2008) has summarized and compared much research done in the area of speech intelligibility. In Darwin (2008) paper he claims paper that according to current research speech is very resilient and humans can understand speech in very extreme conditions. Dialogue will certainly be played at same time as other assets like ambiances and other sound effects and the dialogue may impact how other sound effects are perceived. That is why it is very important to have clear and undistorted dialogue. Quality and intelligibility are not the same thing but they can have an impact on each other. If intelligibility of speech

is sacrificed then the perceived quality may be lowered. If the quality is lower intelligibility may be worse. The similarities between dialogue in games and vocals in music are a few but important, like intelligibility of the lyrics/dialogue, spectral information, and masking of elements. This is important to have in mind when doing sound design for games.

1.5 Audio in games

When producing sound for games there can be multiple types of sounds playing at once. It can be ambiences, music, dialogue, and various sound effects that help the games feel immersive. A game is interactive and non-linear (compared to a movie) and therefore in some way you can control what sounds are supposed to be played. The problem with this is that the player can find themselves in situation that can be hard to understand. If the player for example shoots a gun right next to a non-player character (NPC) that is trying to give the player instructions. Those instructions may go unheard because of the masking from the gunshots. That is one example of why speech intelligibility is important in games (Collins, 2013). As a game developer you can control what a player can do in the game to a certain point but you cannot control certain aspects of the game. If a player decides to stand completely still, the audio still has to contribute to that or not interfere with the player's choice. One way for audio to contribute to the games believability is to be of the highest quality possible.

1.6 Audio quality

Quality is often measured objectively and can vary between people but generally there are some attributes that people value more than others when it comes to audio quality. There are multiple attributes that may describe overall audio quality, intelligibility is only one of them. Others may be clarity, depth, naturalness and spectral balance and they all have an impact on the perceived audio quality. Audio quality may have an impact on the overall quality perception of a game as well. Compression of audio might have a negative impact on quality attributes, but it also provides the necessary disc space for variations.

1.7 Audio compression

Compression of audio started in the late 1980s but to consumers compressed formats were not common until the mid to late 1990s. The lossy format MPEG2 layer 3 (Mp3) increased in popularity because the relative high audio quality and the high compression rate. It has made it feasible to store, stream and transfer large quantities of audio inexpensively. A lossy compression means that the codec reduces the amount of bits used to describe the audio content. The codecs try to identify perceptually irrelevant sounds (like high frequency content) and remove those bits. There are also lossless formats like FLAC, and ALAC that still compress the audio but with no audible consequences. Lossy compression may have audible differences between the lossless original. However the lossless formats have a smaller compression rate and becomes larger than lossy formats. Internet file-sharing has completely changed the way we use data from being limited to physical formats to the point now where almost everything (not just audio) can be streamed directly on demand. It is only a matter of time before all games can be streamed as well (attempts has been made but no success yet) but for now the industry uses discs or space limited physical media.

There have been many tests on compressed audio, for example Liu et. al. (2006) and research done by Beerends and Pocta (2015), that investigate if compression codecs are audible in music. Beerends and Pocta (2015) has established that in higher bit rates (>128 kbit/s) are hard to discriminate from lossless audio. Beerends and Pocta (2015) tested this on music, that may be easier to discriminate compared to dialogue audio due to frequency information, tonal information and transients. But none of the literature explores lossy compressed audio in games even though it's widely used.

In games lossy compression creates more variables that may have an effect on sound quality like; cascade coding, multiple lossy coded stimuli playing at the same time, artifacts from lossy coded audio and masking with other assets. Multiple games today reach around 30 GB in size (The Division by Ubisoft is 32.8 GB, Fallout 4 by Bethesda is 28.8 GB, Star Wars: Battlefront by DICE is 27 GB, Titanfall by Respawn Entertainment is 50 GB). It is hard to say how much of that is audio. The files that usually take up the most space are the PRFMV files (Sims, 2014). The videos are so large because of the high resolution and compared to other visual assets like game textures they can't be calculated and must be read in real time. It is hard to say how much of this disc space is used by audio and it is very few developers that are open with their exact content. If we could find a way to use disc space more efficiently there might be a good thing to use more space for audio.

When doing sound design for a game you also have to take CPU into account. Audio usually requires about 10% of the total CPU usage of a game to run smoothly (Schissler & Manocha, 2011). The audio engine can run better if it uses an effective cache allocation. That means that the engine can store repetitive sounds and ambiences in the cache storage so the system can retrieve information faster (Boer, 2002). The price of CPU power has also dropped so now we can play more advanced games with a more advanced sound design for a lesser price.

In the gaming industry there is a lot of proprietary, secrets and misinformation. Some claim that since the previous generation of consoles (PS3, Xbox360) they have been able to provide lossless audio (Flacy, 2013). It is true that the PS3 can send lossless audio out from its optical or HDMI port but what can we consume that is of lossless audio quality. For sure are no games available today with completely lossless audio. The PS3 is perfectly capable of delivering some HD audio formats like High Fidelity Pure Audio and you can play Wave files from the directory but no games are purely lossless audio.

1.8 Unreal Engine

Another thing that impacts the audio quality in a game is the game engine. Unreal Engine 4 (UE4) is a free open source game engine for independent developers. The source code is developed and shared by Epic Games so developers may build add-ons and it is free for non-commercial use. As a developer you can either choose to develop the game in code (C++) or with a graphical interface. The graphical interface lets you connect nodes much like you connect different hardware with cables but within a digital environment, making it easy to develop complex interactions with little code. The engine has some ways of manipulating sound like reverb, pitch, loop and compression but the ways are limited. The problem

with ready made engines is that the developer is limited in what can be developed by the functionality offered by the engine.

The game developers don't share details of their audio engine and that is why this test is using UE4. UE4 is a popular game engine but not many triple A titles (the major blockbuster game titles) are released that are developed with UE4. Most triple A games are developed by companies with their own game engine. And from a company's point of view it is easy to understand why they won't release details on their engine when the industry is as competitive as the gaming industry.

1.9 Compression of audio in games

This test will focus on the compression part of the audio engine, which allows us to use multiple variations of audio assets but audio quality might be sacrificed. UE4 uses the format Ogg Vorbis to code their audio. By default the compression setting is set to 40/100 which is about 104 kbit/s. If 104 kbit/s is set by default this default presumes that audio of this quality will seem good enough (at least) to players. On Vorbis web page they claim that Ogg Vorbis sound better than Mp3 (Vorbis.com, 2003). This could be based on the claim that people seem to be more critical to the compression when listening to Mp3 then to Ogg Vorbis files (Pocta & Beerends, 2015). People tend to hear the compression when listening to Mp3 files but not as much when listening to Ogg Vorbis files. This generation of gaming platforms (PS4, Xbox One) are able to play lossless formats but all the games developed now use compressed audio. It is known what impact audio compression can have on music but it is hard to predict what people think about compressed dialogue audio in games.

1.10 Purpose

This research will investigate whether players with no special training in audio production can perceive a difference in audio quality in game dialogue while playing. If a pattern can be found that untrained listeners find the default bit rate of 104 kbit/s too low quality or there are no perceived differences between that and higher bit rates an argument can be made that you should raise or lower the default bit rate. This research will test whether it is possible to use lower bit rates or recommend higher bit rates. To answer this question, an experiment will be conducted with an in-game listening test in which subjects will evaluate audio quality.

2. Method

2.1 Listening test

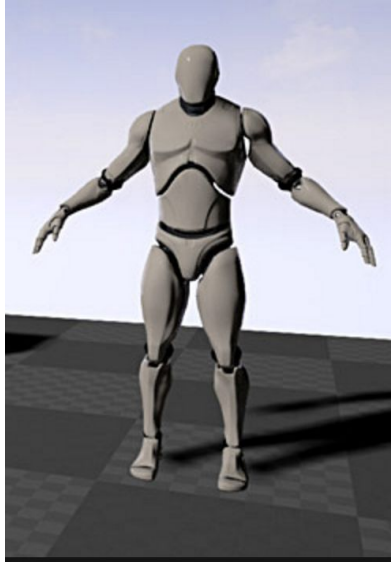
To test if people have a preference when it comes to compression of dialogue in games a listening test was done. The interface to conduct this test was built in the Unreal Engine 4. A level was designed to have similarities to a MUSHRA test (ITU-R BS.1534-3) but embedded in a game. MUSHRA stands for Multiple Stimuli Hidden Reference and Anchor. The anchor in this case will be the stimuli with the lowest bit rate. The anchor is used for the subjects to learn to recognize and analyse lossy coding and its artifacts. This is assumed that the anchor will be heard as poor audio quality so subjects can evaluate higher qualities against it.

The level was designed so much data could be collected in a small amount of time to avoid listening-fatigue. The subjects were asked to choose which two NPCs that sound the best. The test was developed so the subjects would choose two NPCs so that more data could be collected. The data collected is nominal data when subjects are asked to sort the stimuli between two preferred and the four not preferred. There is no order of which the subject will evaluate and choose the two NPCs. The results will be analyzed without internal order taken into account of the two chosen NPCs. The data was then analysed with a chi-squared test to see if the results have a significance and if the null hypothesis can be rejected.

Best sounding dialogue among six choices was the only criteria the subject evaluated. It is a study of preference and if the subjects have a preference of a certain bit rate for dialogue. If a subject prefers a lower bit rate in the experiment, then they are likely to tolerate in-game dialogue compressed at that bit rate. However, preferences could be situational, the subjects may feel that a bit reduced dialogue fits depending on the game.

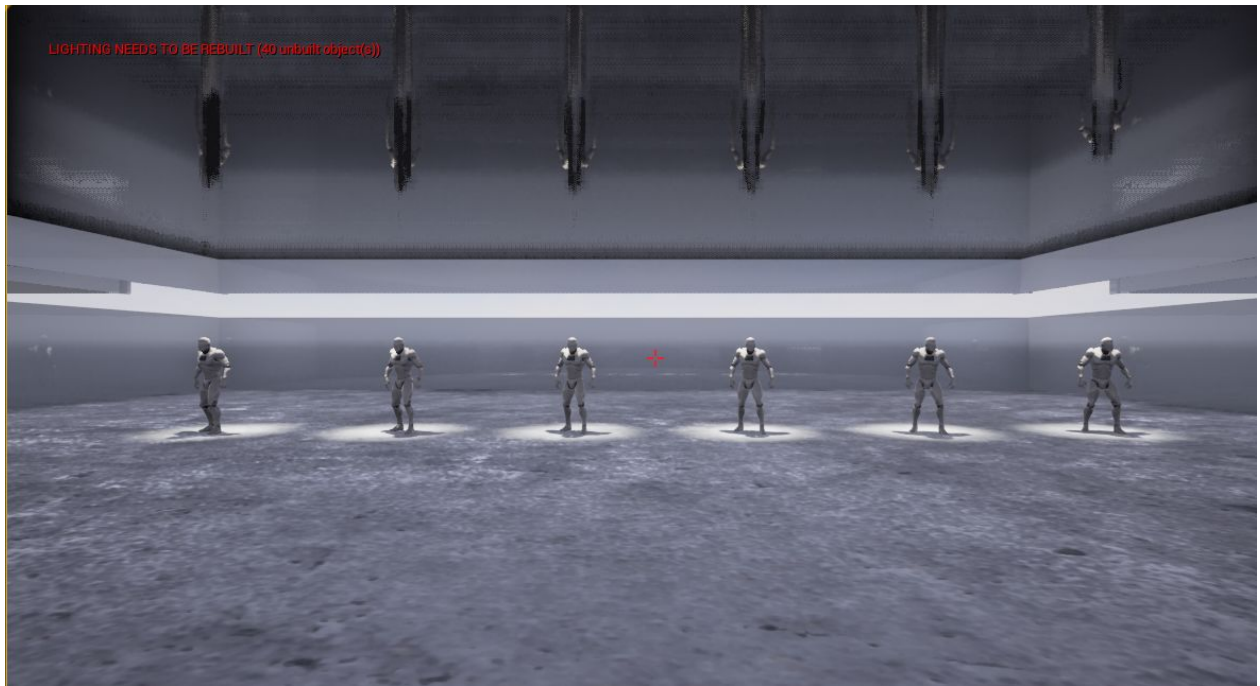
2.2 Development and design

The test was developed in Unreal Engine 4.10.2 which is a game engine software that is very popular for game development. The test design was created so that the subjects would play four levels in which they would choose the two robots or “non-human-characters” they thought sound the best. These characters spoke like humans but looked like robots. The UE4 default mannequin was used for the robot design as seen in *Figure 1*. The mannequin doesn’t have any personal characteristics and fit the test. The aesthetic of the game was futuristic and industrial, just so the NPC-models would fit the scenario. On the levels surfaces concrete and grey standard wall was used. The lights above the NPCs was used so the game would get a more dramatic feel to it. All this was to retain ecological validity. The idea was to have subjects listen in a game environment.



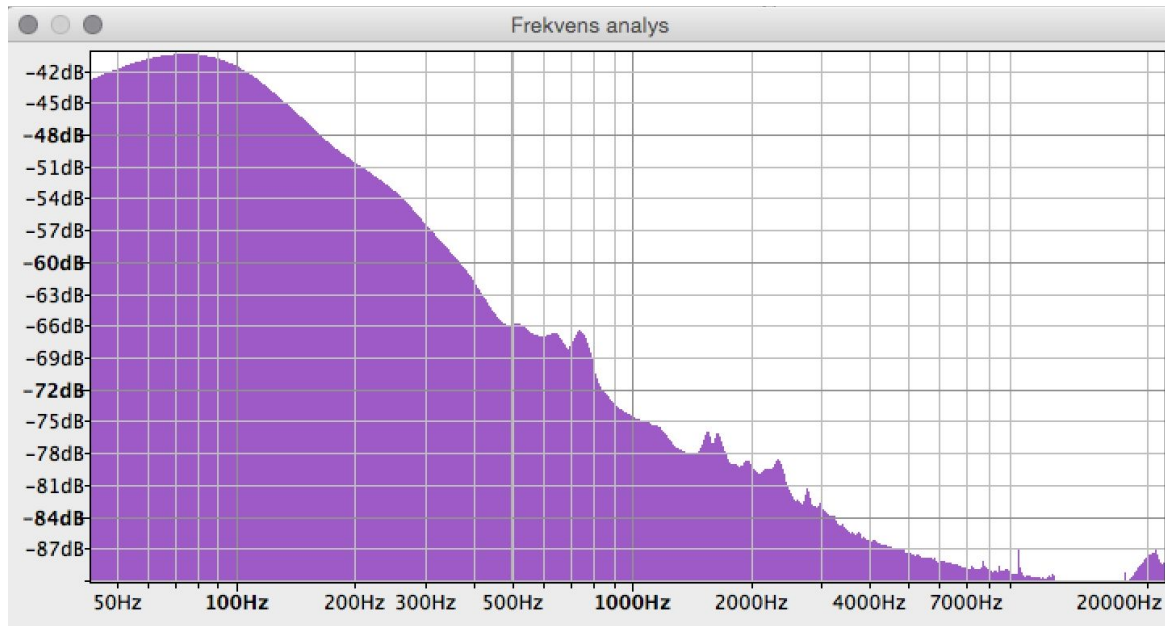
(Figure 1: The mannequin model used for the test)

The subjects play in a first-person point of view and the player is controlled with mouse and keyboard as seen in *Figure 2*. They control the player's movement with keys W, S, A, D and camera-angle or point-of-view with the mouse. This is very common in first-person games and most gamers are used to the controls. The audio stimuli was triggered with trigger-boxes when the player would enter a certain area around the NPC, the audio stimuli would end when the player walked out of that area.



(Figure 2: First screenshot of the game environment)

For increased ecological validity foley and some ambience sounds were added to make the level design feel more natural. Footsteps and cloth foley was recorded with a Sony PCM-M10 with boots on a concrete floor. An ambience sound was used that would not interfere with the dialogue. The ambience was a room tone with a hum with some reverb, a spectral analysis of the ambience can be seen in *Figure 3*.



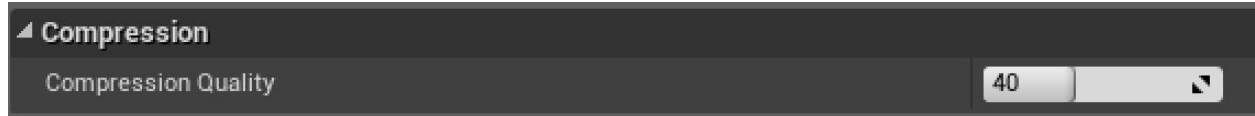
(Figure 3: Spectral analysis of the ambience)

The reverb used was native to UE4 and set a natural level determined by the experimenter that would not be disturbing and would seem applicable to the environment. The ambience and foley assets compression was set to the default compression setting of 40/100. The compression setting of 40/100 for foley and ambience was chosen because this is the default compression rate in UE4.

A system in UE4 was made so the actors will spawn differently for every time the game is started to counteract systematic errors. The randomization in this test is crucial, for example if a subject finds out that all the NPCs to left sounds worse than the once to the right the subjects can detect a pattern. With this system the NPCs spawn randomly across all four levels but to predetermined spawn points. When the players were done with one level they go out of a gate at the back of the level and the player will be instantly teleported to the next level. When the player were done with the fourth level they can teleport back to level one if the subject feel the need to redo one or more levels.

2.2.1 Compression

In UE4 the compression rate of audio was set to a scale of 1-100 where 100 is the least compression and highest quality and 1 is the highest compression rate and the lowest quality. This is set to a bar so you can change the compression setting of each audio asset individually as seen in *Figure 4*.



(Figure 4: When a sound asset in UE4 is clicked a dialogue window will open with this setting with the default set to 40/100)

In UE4 they call it compression quality which is a bit misleading because it is not the quality of the compression it controls, it is the quality of the audio. With this setting the dialogue could have individual compression settings. The compression settings used in the test was 1/100, 20/100, 40/100, 60/100, 80/100 and 100/100, so six different settings could be used and the default of 40/100 would be tested (Epic Games. n.d.). When importing audio in UE4 the compression quality setting is set to 40/100.

The bit rates were determined by exporting the individual audio assets and analyzing them with a ogg analyzer called ogginfo (Amorim R. n.d.). The analyzers results showed that the compression in UE4 was scaled from 45 kbit/s to 288 kbit/s. Multiple audio files were exported, stereo, mono, music, dialogue and the bit rates stayed the same.

2.3 Stimuli

It was important when developing the test that subjects would not have any preferences and tendencies to any NPC. Because the audio is attached to the model, the results had to be tracked in an easy way that would eliminate confusion. There would be 24 different models, one for every six different bit rates, two male and two female versions. To keep track of the choices the subjects made, the NPCs got names and they got the top 24 surnames in America. The names were randomized to the NPCs. The names were chosen by the experimenter to make it easier for the subjects and use the time focusing on the task at hand. The tables with the names and bit rates of the robots found below. The NPC would say his or hers name when the player approaches so the subjects could identify which NPC they talk to.

Table 1 - Level 1

Name	Gender	Bit rate (kbit/s)
Evans	M	45
Wilson	M	72
Thompson	M	104
Watson	M	144
Wright	M	208
Harris	M	288

Table 2 - Level 2

Name	Gender	Bit rate (kbit/s)
Smith	F	45
Jones	F	72
Wood	F	104
Robinson	F	144
Taylor	F	208
Williams	F	288

Table 3 - Level 3

Name	Gender	Bit rate (kbit/s)
King	M	45
White	M	72
Jackson	M	104
Green	M	144
Cooper	M	208
Martin	M	288

Table 4 - Level 4

Name	Gender	Bit rate (kbit/s)
Walker	F	45
Lewis	F	72
Brown	F	104
Davies	F	144
Johnson	F	208
Roberts	F	288

(Table 1, 2, 3, 4 - Tables with the names of the NPCs and bit rates)

Spatialization and attenuation was used on the dialogue to add to the ecological validity. Spatialization is when the audio is panned over the stereo field depending on where the player is facing. Attenuation is when the level of the audio is set to be louder when the player is closer to the source and lowered when you are further away. These are common techniques to use in games so the player still have the sense of freedom even when taking part of a conversation.

2.3.1 Dialogue

The test had to be applicable to modern games and it had to have enough dialogue for the subjects to have time to evaluate the perceived audio quality. To allow for both these requirements, a small backstory was created and a script for the voice actors to perform.

“Sir, Captain [Name] reporting for duty. Our convoy was ambushed in quadrant six at about twenty two hundred hours. We eliminated the hostiles with some casualties. We managed to take three prisoners who had provided us with the coordinates for the enemy's base. I can't wait to get out there and give them hell. Tho I'm obliged to say there has been an error in my system check software so it will be your job to determine if I'm fit to go after them.” - Script for dialogue recording.

The actors recorded eight versions of the part “Sir, Captain [name] reporting for duty...” And one version of the rest of the script. This was to eliminate major differences in the spoken dialogue that may have an

impact on perceived quality. The script was written with many variables in mind. It had to be long enough for the subjects to be able to detect quality differences. Different vowel and consonant sounds may have a different impact on the coded audio. Words like casualties, ambushed, quadrant and coordinates were used to get a variation of sonic elements in the recordings. All these words have either fast consonants that can trigger a transient response in the codec and/or s-sounds that can trigger the codecs high frequency response. These properties and more are important to speech intelligibility (Darwin, 2008). According to Darwin (2008) the sound of speech varies a lot, and with this in mind the script and choice of voice actors were determined.

One male native english speaker and one female native english speaker were used for the dialogue recording. The male was from Australia and the female was from the United States, both had distinct and very clear voices. The recordings were made with a Thuresson CM402, RME Fireface 800, RME Octamic and Logic Pro X in 98 kHz 24 bit. This equipment was chosen due to the fact that many voice-over recordings are done with this or similar equipment. This equipment is equal to the industry standard with a low noise floor and high clarity. The recordings were made in an acoustically treated room. The recorded dialogue was peak-normalized to -3dB and the signal to noise ratio between dialogue audio and ambience was measured to 40 dB. The unreal website states that it is okay to use a sample rate of 22.05 on dialogue (Epic Games n.d.). After some minor editing of the dialogue the stimuli was bounced to 44.1 16 bit wave files. 44.1 kHz 16 Bit is the highest bitrate UE4 can import and process, in this test it is important to limit the effects of cascading that could have been at hand if 22.05 kHz was used.

The test was done in english because most of the games developed are in english. Most of the research in speech intelligibility is also done in english (Darwin 2008). Most Swedish gamers play games in English. So, this was not considered to be a complicating issue in the experiment.

2.4. Listening conditions

The subjects played the game on a Macbook Pro with headphones (Beyerdynamic DT250). Goodwin (2009) say that most gamers use headphones when playing games. Shobben (2004) says that people are more critical to hearing artifacts of lossy compression while using headphones. The test would be done in either acoustically treated rooms or environments where the subjects usually play games in quiet conditions.

2.5 Pilot study

A pilot study was conducted to determine if the test was good enough and if trained listeners could hear a difference between the stimuli. Four trained listeners did the test with the same planned conditions that the subjects would have. After the test was done questions we're asked to determine if there was something in the test they found hard, easy and/or disruptive. After the four subjects did the test, small changes were made to the final test. Names above each NPC were added as seen in *Figure 5*. The names were added so the subjects could easily identify which NPCs they preferred. The reverb was turned down so it was almost inaudible and a bug in the way the system played the ambience was fixed. In the pilot

study one of the subjects thought the ambience sounded strange and out of phase. When developing the game two ambience nodes was accidentally placed in the world. Thus creating two identical sounds and when they don't spawn at the exact same time they become out of phase. All four subjects in the pilot study chose the highest two bit rates for all the levels.



(Figure 5 - Screenshot of the final game with the names of the NPC included)

2.6 Subjects

The subjects for the main test were 24 untrained listeners with gaming experience. The levels of gaming experience varies from subjects who play mobile games and sometimes play on console or PC to serious gamers with >10 hours/week. For this test untrained listeners was used because that represents a majority of gamers. Trained listeners may hear artifacts and can distinguished compressed audio from lossless audio, this test will not focus on that demographic. The subjects were between 21-29 years old from Sweden.

No more attributes or characteristics were asked to take note of except which NPCs the subjects thought sounded the best. One subject did the test in 30 minutes but the rest of the subjects did the test between 10-20 minutes. The tests that were done in Piteå in an acoustically treated and sound proofed room. The testes done in Stockholm were done in the subjects apartments in their living rooms in quiet conditions. The subjects were free to set the volume themselves but recommended to keep the volume at the same level during the entirety of the test.

2.7 Questionnaire

A questionnaire in english was made that the subjects would do before the test was made. The questions were age, english comprehension level, and the subjects gaming experience. The questions for the subjects english comprehension level and gaming experience were multiple choice questions. The subjects would check boxes for how many hours a week they play games from <1 hour / week to >10 hour / week. The subjects would check the box that represents their english comprehension level graded from “I do not speak any english” to “I speak and understand English completely fluently.” On the next page there was a answer sheet for the subjects preferred stimuli and on the third page room for additional comments. The full questionnaire can be found under appendixes. An written version of instructions in english was provided in the answer sheet. The instructions was:

“In this test you will be asked to evaluate audio quality of speech. The test will be done with an in-game environment. You will be playing four levels, each with a set of six robots. Each of the six robots has a different voice, they may seem similar but they are different. When you approach the robot the robot will start to talk. When you have talked to the robots you will decide which **two** robots sound the best. When you are done with one level you can teleport to the next level by going out through the gate. This test will take about 20 minutes but please take your time. Answer sheet is on the next page. “

The questions about gaming experience and english comprehension was added to be able to detect patterns or influences in subjects responses. If an answer pattern could be found in subjects with lower english comprehension level an argument can be made that english comprehension has an impact on overall quality assessment of dialogue. The same can be said for gaming experience.

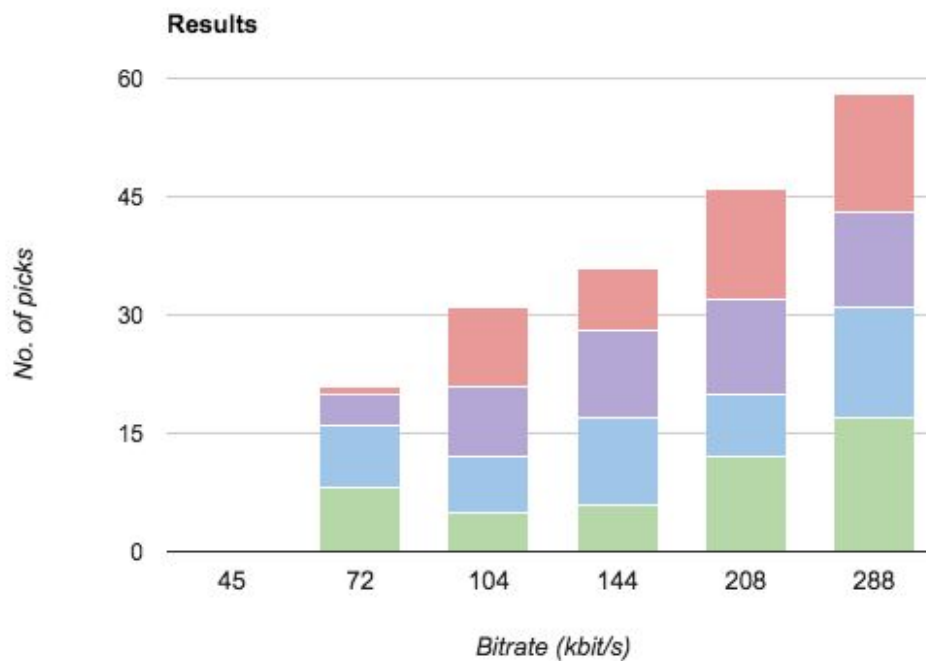
3. Results

In this section the results will be presented with different figures and tables. In *Table 5* the overall results are presented divided into the bit rates tested. In *Figure 7* and *Table 6*, *7*, *8*, and *9* the results are divided into the four different levels. In *Figure 8* and *Figure 9* the results are divided into genders, *Figure 8* presents the male levels results and *Figure 9* presents the female levels results. The subjects were asked to choose which two NPCs that sounded the best and the results of the 24 subjects can be seen in this section. The results are divided into different tables and figures for easier understanding.

Table 5

Bit rate (kbit/s)	No. of picks
45	0
72	21
104	31
144	36
208	46
288	58

(Table 5 - The subjects have picked more NPCs with higher bit rates. Each subject would choose two NPCs and this is the combined results from all the subjects)



(Figure 7 - Results divided into individual levels and bit rates, green - level one, blue - level two, purple - level three, red - level four)

In *Figure 7* we can see that most of the subjects preferred higher bit rates than the lower alternatives. In the results there are no clear threshold for when the subjects stop to hear an increase in perceived audio quality. A threshold would look like the curve would flatten after a certain point, in *Figure 7* we can't detect such a pattern occurring.

Table 6 - Level 1

Name	Gender	Bit rate (kbit/s)	No. of picks
Evans	M	45	
Wilson	M	72	8
Thompson	M	104	5
Watson	M	144	6
Wright	M	208	12
Harris	M	288	17

Table 7 - Level 2

Name	Gender	Bit rate (kbit/s)	No. of picks
Smith	F	45	
Jones	F	72	4
Wood	F	104	9
Robinson	F	144	11
Taylor	F	208	12
Williams	F	288	12

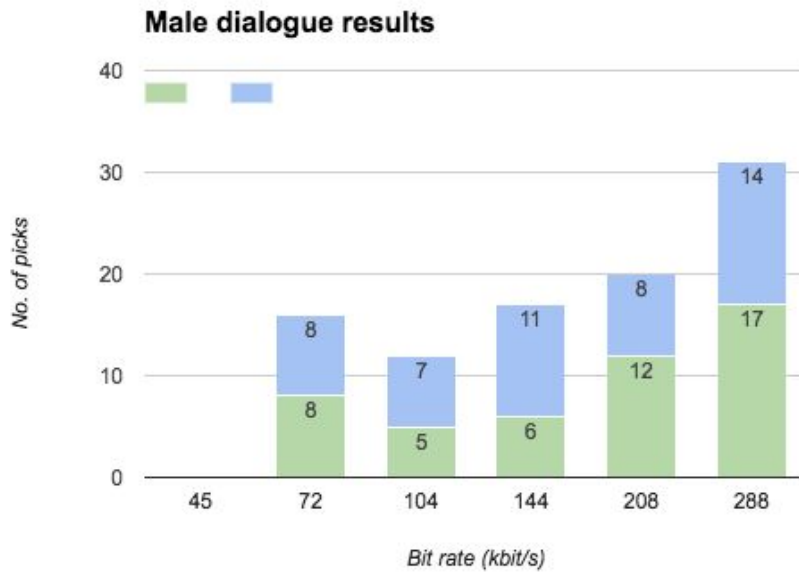
Table 8 - Level 3

Name	Gender	Bit rate (kbit/s)	No. of picks
King	M	45	
White	M	72	8
Jackson	M	104	7
Green	M	144	11
Cooper	M	208	8
Martin	M	288	14

Table 9 - Level 4

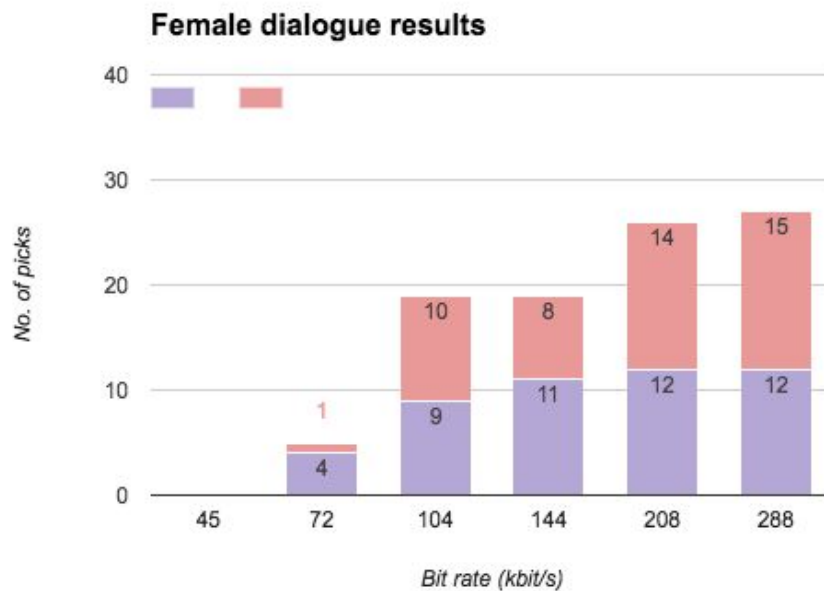
Name	Gender	Bit rate (kbit/s)	No. of picks
Walker	F	45	
Lewis	F	72	1
Brown	F	104	10
Davies	F	144	8
Johnson	F	208	14
Roberts	F	288	15

(Table 6, 7, 8, 9 - Tables with the results divided into levels)



(Figure 8 - The results of the two levels with male dialogue, green - level 1, blue - level 2)

In level one and two, the two levels with the male dialogue results pooled we can see that it is harder for the subjects to discriminate higher from lower bit rates (*figure 8*). At least (on visual inspection) bit rates 72 kbit/s to 208 kbit/s. The trend is that more people prefer the higher bit rates and with these bit rates we see no tendencies of a minimal threshold bit rate for perceived quality. A bit rate that the subjects stop hearing a difference, or a bit rate which above there seem to be no preference.



(Figure 9 - The results of the two levels with female dialogue, purple - level 3, red - level 4)

In *figure 9* we can see that the subjects have a slightly higher threshold for discriminating bit rates. In the female dialogue the subjects seem to prefer higher bitrates but very evenly spread out across four bit rates. If you compare *figure 8* and *9* you can see that there is a more apparent spread to the levels with the male voices than to the levels with the female voices.

3.1 Comments from subjects

The task - Multiple subjects (6) commented on the anchor stimuli. That it was easy to find which of the six stimuli that sounded the worst but harder to find which stimuli that sounded the best. Two subjects commented that it was fairly easy to find the first of the two stimuli they thought sounded the best but found it harder to find the second one.

Reverb - One subject commented on how the subjects thought the reverb changed the quality. The subject said that they chose based on perceived reverb quality.

Memorizing - One subject commented on the difficulty to memorizing how the stimuli sounded between them. The subject suggested a shorter dialogue or to use some way to jump to different parts in dialogue.

The sound - One subject commented on the s-sounds and the perceived difference between them. The subject said that he thought that on the lower quality dialogue the s-sounds became harsh.

Male dialogue compared to female dialogue - One subject commented that the subject thought it was easier to discriminate higher from lower bit rates on the female dialogue.

The comments and the full result sheet can be found in appendix 2. The comment section did not specify which language the subjects should write the comments in so many of the comments are in Swedish.

4. Analysis

4.1 Comments

One subject chose the stimuli that had the best sounding reverb that sounded natural. However, the only parameter changed in between NPCs are the bit rate of the dialogue. If the subjects thought that the reverb sounded the best on one NPC then that means that the reverb sounded the best with that compression setting. This subject chose two 72 kbit/s, three 104 kbit/s, one 144 kbit/s and two 288 kbit/s.

One subject commented on the s-sounds, and that might mean the subject may have found a fixed type of sound that the subject could evaluate the stimuli through this sound alone. With that the subject may have been able to discriminate the stimuli from each other. When looking through that subjects test there doesn't seem to be as clear as the subjects comment. The subject have chosen one 72 kbit/s, three 104 kbit/s, two 144 kbit/s and two 288 kbit/s. If this subject could clearly identify better sounding dialogue there should be more consistency in his or her choice. Maybe the compression will have an impact on the s-sounds and reverb but it doesn't seem to have to improve this subject's ability to discriminate higher from lower bit rates.

The test was developed so the subjects would only determine audio quality with the only criterion that “what sounds best”. The test did not specify this but it can be assumed that if the only changed parameter is compression if the other attributes changes it may have a connection to compression.

The subject that commented that it was easier to pick the female dialogue picked one 144 kbit/s, three 208 kbit/s and four 288 kbit/s. The lowest bit rate picked (144 kbit/s) was the male voice, but the subject in general could discriminate higher from lower bit rates.

4.2 Quantitative data

The results show that untrained listeners for the most part can discriminate higher bit rates over lower bit rates (*figure 7, tables 7 to 9*). The results show no sign of a threshold for preference being reached, then the curve would have been flattened when reaching a critical bit rate (*figure 7*). Berg et, al. (2013) says that the threshold for quality detection is around 300 kbit/s in DAB+. Of the subjects answers, 21% of the answers are bit rates of 104 (UE4 default) or lower. We can see there is a slightly bigger spread on the male voices compared to the female voices (*figure 8 and 9*).

4.3 Chi-Squared test

The results were analysed with a chi-squared test like Andersen (2011) recommends, with the help from a calculator created by Preacher (n.d.). The analysis was using an alpha value of 0.05 and when the subjects had six stimuli to choose from the analysis will have five degrees of freedom the critical value is 11.070. If the chi-squared value exceeds 11.070 then the null hypothesis can be rejected. The chi-squared value for this tests were calculated to 63.563 thus exceeding 11.070. With a chi-squared test we can say that the results are not based in chance and there is a significant difference between the observed data and the

expected results. Even if changing the alpha value of 0.001 the measured chi-squared value exceeds the critical value. When analysing both the male results and the female results separately there is still a significant difference between the observed data and the expected data. The male chi-squared value is 32.125 and the female is 38.5. We can see there is more of a spread in the male tests but the chi-squared score is still significant which means the answers are not random. All calculated values were manually controlled to confirm the results.

5. Discussion

5.1 Reflections on the results

This test confirms that untrained listeners can detect differences in bit rates when listening to dialogue in games. When only 21% of all the picks were on the critical UE4 defaults 104 kbit/s or lower it is clear that UE should raise the default bit rate for dialogue (*figure 7*). This can vary depending on games, but dialogue heavy and story driven games would benefit from a raised default bit rate.

When comparing the levels with the male dialogue and the levels with the female dialogue we can see that it's slightly harder for subjects to discriminate lossy coding with the male dialogue (*figure 8 and 9*). That may be because of the actors Australian accent that may mask some of the artifacts or that the subjects are more used to hearing the female U.S accent. It is true that most of the games today have more american voice actors, could that have an impact on discriminating lossy compression. When listening to the dialogue the male and the female are not that far apart when it comes to fundamental frequency, the female dialogue is a little bit higher than the male. This was checked in the original Logic Pro X project with the plug-in Multimeter by comparing words. It may be so that the higher voices enhance the artifacts created by lossy compression.

Preferred is not equal in all cases to the best sounding. The subject can prefer a stimuli that have lower sound quality but fits the scenario and becomes believable. That is important to take into account when doing similar tests. It was only asked of the subjects to choose the best sounding NPCs, the results might have differed if they were asked which ones they preferred.

5.2 Critique of the method

The scenario could have been designed in multiple ways, AB, ABX or a grading scale. To get the most streamlined test experience this test was developed where the subjects would choose 2/6 NPCs which they thought sounded the best. It has similarities with a MUSHRA test but also with a two comparison test because the test examines if the default bit rate is good enough. If there would be a random result we should have seen around 32 picks on every bitrate, even when using pairs like this test did. This method was chosen also because it was implementable in an in-game environment with little coding effort.

5.3 Subjects

This test was done with subjects that had no special training in audio. That was because most of gamers do not have special training in audio and it was crucial to get data that is representable for the population. The subjects were all swedish and had english as a second language. 16 out of the 24 subjects claimed they speak and understand english completely fluent. And the rest of the subjects are divided between "I can speak and understand reasonably well and can use basic tenses but have problems with more complex grammar and vocabulary" and "I speak and understand well but still make mistakes and fail to make myself understood occasionally". No apparent patterns was detected between the subjects english comprehension level and their perceived quality assessment of english dialogue.

5.4 Graphics and audio influence

The aesthetic of the game is futuristic and industrial and that might have had an influence on the perceived quality of the dialogue. The subjects might have expected a certain sound design, that might influence our quality perception. The mannequin model was chosen because there would be no special human connection and the problem with lipsync and visual preference would not be a problem. It may have been so that some of the subjects thought the model sounded more believable with dialogue in a lower bitrate. It may have also been the other way around, that without lipsync and advanced character design the subject could truly focus on the dialogue and be more discriminating. It was important to create an environment that was believable and still non-distracting and this level- and character design achieved that.

5.5 Ecological validity

In this test the ecological validity was of high importance so questions like, how gamers play their games and how do they usually listen had to be answered. Goodwin (2009) claim that most gamers listen in headphones and fortunately Schobben (2004) claims that people are more critical to artifacts when listening in headphones. The laptop was used simply because of its mobility, the test could be moved to any location. That helped to gather subjects from both Stockholm and Piteå, most of the subjects are students at LTU Campus Piteå. The tests done in Stockholm were not tested in an acoustically treated environment, but when studying the results you can see no difference between the Stockholm and Piteå subjects. The tests done in Stockholm were done in environments where the subjects usually play games.

It would be easy to do a similar test without the game, but then it would lack ecological validity. These results can be a direct consequence from the subjects having a higher level of interactivity. When the subjects can move around freely and choose which order to listen to the stimuli they may focus on other things than just the audio. The results show that they still can discriminate higher from lower bit rates in dialogue audio.

5.6 Expected results

It is assumed in this test that people could discriminate the stimuli and find the ones with the higher bit rates better sounding. Gamers have for the most part grown up with poor audio quality and it could be possible for that to reflect in these results. If the test would have been done on experienced listeners I think the results might have a steeper slope and more would pick the higher bit rates.

5.7 Dialogue assessment

The test is only done with dialogue because there is so much dialogue in certain games. If sound designers were to choose one type of asset to compress, it makes sense to compress dialogue. Since there is so much dialogue in games, the quality of dialogue has to contribute to any overall audio quality assessment in games. That is because dialogue has so much variations and it is easy to know what dialogue is supposed to sound like. It may be difficult to determine variations in quality when examining an explosion or foley

because it can sound so different. Some games are often meant to simulate real world interactions which means both content and quality. Compression artifacts does not occur in real life but at what level do we find the compression annoying or immersion breaking? This test does not answer that but it seems that people think dialogue coded in higher bit rates sound better.

5.8 Optimizing Compression

With this test in mind it would seem to be beneficial for UE to make use of the full capabilities of the Ogg format that can reach up to 500 kbit/s (Vorbis.com, n.d.).

The idea of dynamic compression is to use the most effective way of compressing audio so the compression becomes inaudible. If you could analyze masking sound you may find that the ambience may be redundant and may lower its bit rate to a degree. If you could make multiple analysis a system may be developed to for an effective use of bits. This is the basic idea behind psychoacoustic compression (e.g., MP3). You may find that dialogue in a cut scene or PRFMV will need higher bit rates than dialogue/shouts in a battlefield so you could develop a way to use dynamic compression. If a system can be developed that changes the bit rates of all the assets depending on what level masking occurs the gamers experience may be better. This would be easy for non audio experts to work with then it will just be one bar with audio quality level. If the bar is lowered then more compression will be made and more compromises and sacrifices will be made. If you raise the bar the compression ratio may become lower and you get to use more bits for audio. This would be nice if it would be implemented alongside the dynamic compression so the games sound have an overall audio quality. Just like audio codecs (Mp3, Ogg Vorbis etc.) that take into account the different attributes of the sound and then removes the least noticeable bits. The dynamic compression in games will do this but with multiple assets and it will work in real time because of a game's non-linear pattern. If much research is done on masking, perceived quality and compression in games multiple factors may come up that may help to develop this system.

5.9 Conclusion

In this test we can see that gamers find higher bit rates better sounding than UE4 default bit rate of 104 kbit/s in dialogue audio in games (*figure 7*). This test was done in an in game environment that was done on 24 untrained subjects listening in headphones. The audio in games is still compressed even though storage has become cheaper and computers have gotten faster. With these results an argument can be made that game developers should take in consideration what and how they compress their dialogue, especially regarding bit rate. The subjects seem to be a bit more critical towards the stimuli with the female voice (*figure 8 and 9*). That may be caused by the timbre in the voice but it may also be the actress or actors difference in accent. This research could start a trend to use higher bit rates for audio in games and that may mean that the gaming experience can be improved. If more combinations of sounds and bit rates were to be tested a more thorough answer may be given if to what level people care about audio quality in games.

5.10 Further research

There are many ways you can further this study but the next logical step would be to examine how sound effects and ambience compression impact overall audio quality. What if lossy coded ambience have a severe impact on every other sound in the soundscape? If raising the ambience compression level would mean an overall increased perception of higher audio quality. The next step after that would be how much cascade sound design has an impact. Maybe 288 kbit/s dialogue sound the best on itself but if you use a 104 kbit/s ambience the overall quality may be sacrificed. You may find that a much higher perceived audio quality may be accomplished by using 144 or 208 kbit/s on both of the assets. That may save disc space and be an effective way to implement. If the compression settings are the same on all assets it may be that some artifacts are masked or enhanced. If different bit rates are used on different assets, the artifacts that may occur can seem to stick out and become more audible. These are only theories about the next step to come to a compromise that both sounds good and use a high compression setting. You can also evaluate different codecs and engines. Because this is a very secretive matter between competing companies it will be hard to follow through but it would be worth testing.

Acknowledgements

I want to thank following people for supporting and contributing to my work.

Nyssim Lefford, Assistant Professor at Luleå University of Technology, for being my supervisor and her encouragement throughout this project.

Kathryn Cederborg and Christopher Pantelidis for excellent voice-acting.

Johan Ehn, Camilla Näsström, Gustaf Terneborg and Johan Westling for participating in my pilot-study.

Erica Engborg for helping me with finding and gathering information in an area of her expertise.

References

- Alten S.R. (2014). *Audio in Media: 10th ed.* Published at Cengage Learning, Wadsworth USA.
- Amorim R. (n.d.). *OggInfo*. Retrieved February 23, 2016, from <http://www.rarewares.org/ogg-tools.php>
- Andersen P. (2011, November 13). *Chi-Squared Test*. Retrieved March 10, 2016, from <https://www.youtube.com/watch?v=WXPBoFDqNVk>
- Berg, J., Bustad, C., Jonsson, L. Mossberg, L. Nyberg, D. (October 2013). Perceived Audio Quality of Realistic FM and DAB+ Radio Broadcasting Systems. Published in *Journal of the Audio Engineering Society*, Vol. 61, No. 10
- Boer J. R. (1 September 2002). *Game Audio Programming*. Published at Cengage Learning, Wadsworth USA.
- Chandler, H. M. (2009) *The Game Production Handbook: 2nd ed.* Jones and Bartlett Publishers.
- Chi-Min Liu, Han-Wen Hsu, Chung-Han Yang, Shou-Hung Tang, Kan-Chun Lee, Yung-Cheng Yang, Chia-Ming Chang and Wen-Chieh Lee. (2006, October) Compression Artifacts in Perceptual Audio Coding. Paper presented at the 121th Convention of the Audio Engineering Society, San Francisco, USA.
- Collins, K. (2013) *Playing with Sound : A Theory of Interacting with Sound and Music in Video Games*. Cambridge, MA, USA: MIT Press, 2013.
- Darwin, C.J. (12 March 2008). *Listening to speech in the presence of other sounds*. Published by The Royal Society Publishing, Philosophical Transactions B. London, England.
- Epic Games. (n.d.). *Audio Files*. Retrieved February 15, 2016, from <https://docs.unrealengine.com/latest/INT/Engine/Audio/WAV/index.html>
- Flacy, M. (2013, September 16). *Demystifying the audio settings on your PlayStation 3*. Retrieved March 27, 2016, from <http://www.digitaltrends.com/home-theater/demystifying-the-audio-settings-on-your-playstation-3/>
- Goodwin S. N., Central Technology Department, Codemasters Software Company, UK (February 1, 2009). How Players Listen. Paper presented at the 35th AES International Conference: Audio for Games, London.
- ITU-R BS.1534-3. *Method for the subjective assessment of intermediate quality levels of coding systems*. (2015, October 23). Retrieved March 13, 2016, from <http://www.itu.int/rec/R-REC-BS.1534-3-201510-I>

Martin, M. (2015, September 3). *Fallout 4 has more lines of dialog than Fallout 3 and Skyrim Combined*. Retrieved March 23, 2016, from <http://www.ign.com/articles/2015/09/04/fallout-4-has-more-lines-of-dialog-than-fallout-3-and-skyrim-combined>

McDonald, J.H. (2014). *Handbook of Biological Statistics (3rd ed.)*. Sparky House Publishing, Baltimore, Maryland. Retrieved February 9nd 2016 at <http://www.biostathandbook.com/chigof.html>

Vorbis.com (2003). *Ogg Vorbis FAQ*. Retrieved February 2nd 2016 at <http://www.vorbis.com/faq/#lossy>

Pocta P., Beerends J. G.. (2015, September). Subjective and Objective Assessment of Perceived Audio Quality of Current Digital Audio Broadcasting Systems and Web-Casting Applications. Paper published in *IEEE Transactions On Broadcasting*, Vol. 61, No. 3.

Preacher, K. J. (n.d.). *An interactive calculation tool for chi-square tests of goodness of fit and independence*. Retrieved March 09, 2016, from <http://www.quantpsy.org/chisq/chisq.htm>

Schissler, C., & Manocha, D. (2011). GSound: Interactive Sound Propagation for Games. Presented at the AES 41st International Conference, Audio for Games, London, UK.

Schobben D., Van De Par S.. (October, 2004). The effect of room acoustics on MP3 audio quality evaluation. Paper presented 117th International Audio Engineering Society Convention San Francisco, USA.

Sims, D. (2014, October 17). *AAA game-file sizes are getting out of hand*. Retrieved March 13, 2016, from <http://venturebeat.com/community/2014/10/17/aaa-game-file-sizes-are-getting-out-of-hand/>

Suspension of disbelief. (n.d.). Dictionary.com's 21st Century Lexicon. Retrieved May 17, 2016 from Dictionary.com website <http://www.dictionary.com/browse/suspension-of-disbelief>

The Entertainment Software Association (ESA). *2015 Essential Facts About The Computer And Video Game Industry*. (n.d.). Retrieved March 23, 2016, from <http://www.thesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf>

Appendix 1

Questionnaire

Bachelor thesis in Audio Engineering by Anton Ahlberg

Age:

Your english comprehension level:

- ☐ I do not speak any English.
- ☐ I can communicate simply and understand in familiar situations but only with some difficulty.
- ☐ I can speak and understand reasonably well and can use basic tenses but have problems with more complex grammar and vocabulary.
- ☐ I speak and understand well but still make mistakes and fail to make myself understood occasionally.
- ☐ I speak and understand English completely fluent.

Do you consider yourself a gamer?

- ☐ Yes
- ☐ No

If yes, how many hours / week do you play games?

- ☐ <1 hour /week
- ☐ 1-3 hours /week
- ☐ 3-5 hours /week
- ☐ 5-8 hours /week
- ☐ 8-10 hours /week
- ☐ >10 hours /week

Task description:

In this test you will be asked to evaluate audio quality of speech. The test will be done with an in-game environment. You will be playing four levels, each with a set of six robots. Each of the six robots has a different voice, they may seem similar but they are different. When you approach the robot the robot will start to talk. When you have talked to the robots you will decide which **two** robots sound the best.

When you are done with one level you can teleport to the next level by going out through the gate. This test will take about 20 minutes but please take your time. Answer sheet is on the next page.

Check the boxes of the two robots you think sound the best.

Level 1

- ☐ Wright
- ☐ Wilson
- ☐ Harris
- ☐ Evans
- ☐ Thompson
- ☐ Watson

Level 2

- ☐ White
- ☐ Green
- ☐ Martin
- ☐ Cooper
- ☐ Jackson
- ☐ King

Level 3

- ☐ Williams
- ☐ Robinson
- ☐ Jones
- ☐ Wood
- ☐ Smith
- ☐ Taylor

Level 4

- ☐ Brown
- ☐ Walker
- ☐ Lewis
- ☐ Johnson
- ☐ Roberts
- ☐ Davies

Thoughts and comments:

Thank you for participating!

Appendix 2

Table with results

Subject	Ivl 1	Ivl1B	Ivl2	Ivl2B	Ivl3	Ivl3B	Ivl4	Ivl4B	Age	English	Gamer	Hours	Comments
1	Wilson	Harris	Martin	Cooper	Williams	Jones	Brown	Lewis	25	5 / 5	y	4 / 6	
2	Wright	Harris	White	Jackson	Jones	Wood	Johnson	Davies	26	5 / 5	y	5 / 6	"Generellt pratade kvinnan för fort. De "burkigare" rösterna fungerar inte alls, generellt de som stod stör.
3	Wilson	Thompson	Green	Martin	Wood	Taylor	Johnson	Roberts	23	5 / 4	y	6 / 6	"Finner det jobbigt överlag med starka s-ljud i spel, valde utifrån det."
4	Wilson	Thompson	White	Martin	Robinson	Wood	Brown	Roberts	22	4 / 5	n	1 / 6	"Jag fick känslan av att jag gillade de röster bäst som fick avståndet till den jag pratade med att verka mest naturligt. Tror är att min egen uppfattning om hur rumsklngen borde låta fick mig att välja som jag gjorde."
5	Harris	Watson	Martin	Cooper	Robinson	Jones	Johnson	Brown	22	5 / 5	y	6 / 6	
6	Wright	Harris	Martin	Jackson	Williams	Taylor	Roberts	Davies	21	5 / 5	y	6 / 6	
7	Wright	Harris	White	Martin	Williams	Taylor	Brown	Johnson	21	3 / 5	n	1 / 6	"En av robotarna var alltid betydligt mycket sämre än de andra. Men i övrigt var det svårt att avgöra."
8	Wright	Watson	Green	Martin	Williams	Taylor	Johnson	Roberts	23	5 / 5	n	1 / 6	
9	Wright	Harris	Green	Jackson	Robinson	Wood	Brown	Davies	27	5 / 5	y	4 / 6	
10	Harris	Watson	Martin	Jackson	Jones	Taylor	Johnson	Roberts	22	4 / 5	y	2 / 6	" Det fanns en tydlig närhet i de jag valde, de kändes mest realistiskt."
11	Harris	Watson	Martin	Cooper	Williams	Taylor	Johnson	Roberts	22	3 / 5	y	2 / 6	
12	Wright	Watson	Martin	Cooper	Williams	Robinson	Brown	Roberts	23	5 / 5	y	2 / 6	"Det var svårt att hitta de två som var bäst då jag ibland bara tyckte om en."
13	Wright	Harris	Green	Cooper	Williams	Taylor	Johnson	Roberts	24	5 / 5	y	5 / 6	"lättare bland de kvinnliga rösterna"
14	Wright	Harris	Green	Cooper	Williams	Robinson	Johnson	Roberts	23	5 / 5	y	6 / 6	"finding the worst was easy, but finding the voice with the highest quality was very hard".
15	Wilson	Harris	Green	Jackson	Williams	Wood	Roberts	Davies	25	5 / 5	y	4 / 6	
16	Wilson	Harris	White	Martin	Williams	Taylor	Brown	Johnson	29	5 / 5	y	3 / 6	"jag hittade den sämsta fort, de bästa var svårt."
17	Wilson	Thompson	White	Green	Wood	Taylor	Roberts	Davies	27	3 / 5	n	6 / 6	
18	Wilson	Harris	White	Jackson	Wood	Taylor	Johnson	Roberts	28	4 / 5	n	2 / 6	"Svårt, tyckte de flesta lät väldigt lika. Men det var alltid en som var sämre"
19	Wright	Harris	green	Martin	Williams	Robinson	Brown	Roberts	27	5 / 5	y	3 / 6	"Tyckte killen pratade lite ottydligt, men alla banor var svåra"
20	Wilson	Thompson	White	Jackson	Robinson	Wood	Brown	Roberts	21	3 / 5	y	5 / 6	
21	Wright	Harris	Green	Martin	Williams	Robinson	Johnson	Davies	21	5 / 5	n	5 / 6	
22	Wright	Harris	Green	Cooper	Wood	Taylor	Johnson	Roberts	21	5 / 5	y	1 / 6	"It was easy to find the worst voice."
23	Wright	Harris	Martin	Cooper	Robinson	Taylor	Johnson	Davies	23	5 / 5	n	3 / 6	
24	Thompson	Watson	White	Green	Robinson	Jones	Brown	Davies	26	3 / 5	n	2 / 6	