TECHNISCHE UNIVERSITÄT
ILMENAU

FAKULTÄT ELEKTROTECHNIK UND INFORMATIONSTECHNIK

MEDIENINGENIEURWISSENSCHAFTEN

MRSP Project

Perceptual Similarity between different Audio Stimuli

Group 2

Project Report

July 31, 2024

Contents

1 Introduction 3

1.1 Problem Statement . . . 3

1.2 Solution Approach . . . 3

2 Tasks 3

2.1 Preparing audio data . . . 3

2.2 Compute the Perceptual Loss between audio for each group . . . 4

2.3 MUSHRA: Subjective measures . . . 4

2.4 Listening experiment management . . . 4

2.5 Report and Coordination . . . 4

3 Task 1 6

3.1 Download of audio . . . 6

3.2 Audacity settings for synthesizing . . . 7

4 Task 2 9

4.1 Brief overview . . . 9

4.1 Algorithm and flowchart . . . 9

4.2 Results . . . 10

4.3 Discussion and Conclusion . . . 12

4.3.1 Genre-Specific Observations . . . 12

4.3.2 Factors Influencing Perceptual Loss . . . 13

4.3.3 Conclusion . . . 14

5 Task 3 15

5.1 Introduction and Objectives . . . 15

5.2 Stimuli . . . 15

5.3 Configuration Steps . . . 15

5.4 Screen Captures . . . 17

5.4.1 Test Creation . . . 17

5.4.2 Test Performance . . . 18

6 Task 4 19

6.1 MUSHRA Listening Test Conduction . . . 19

7 Task 5 20

7.1 Description . . . 20

7.2 Indentifying Outliers . . . 20

7.2.1 Detection and Removal . . . 20

7.2.2 Post Outlier Removal Graph . . . 20

7.3 Insights and Conclusion . . . 20

7.3.1 Acoustic Environment Preferences . . . 21

1 Introduction

1.1 Problem Statement

The main aim of the seminar project is to find out the perceptual similarities between different audio stimuli by calculating

- objective measure: Perceptual Loss Function
- subjective measure: MUSHRA listening experiment
- submission of a comprehensive report with both the findings

1.2 Solution Approach

The project is split into five tasks with five people 1 assigned one task each. Here is the table that describes who is assigned to which task.

Table 1: Group 2

Matriculation Number	email	Task #
65821	azmat-sultan.awan@tu-ilmenau.de	Task 1
66451	bharani.jayakumar@tu-ilmenau.de	Task 2
66512	umair.khalid@tu-ilmenau.de	Task 3
66160	tarun-devidas.ramani@tu-ilmenau.de	Task 4
64365	suprabha.ghosh@tu-ilmenau.de	Task 5

2 Tasks

2.1 Preparing audio data

- Prepare five dry audios with 24-bit rate and sample rate of either 44100 Hz or 48000 Hz. These audios should have distinct spectral characteristics which means select five different audio types such as:
  - Drum
  - Speech
  - Piano Music
  - Mix Instrumental Music
  - Trumpet (These are just examples, feel free to choose)

- From these five audios, generate three versions of each with:
  - Small Office Environment
  - Opera Hall
  - Reverb hall with reverberation time at least 1.5 sec
- In this way you have 20 audio files with 5 groups (each group containing 4 versions of an audio)

2.2 Compute the Perceptual Loss between audio for each group

- You need to use the GIT repository for this purpose, which is available at: https://github.com/TUIlmenauAMS/PsychoacousticLoss

2.3 MUSHRA: Subjective measures

- Prepare the MUSHRA test which will be used to perform a listening experiment (hand over the complete functional MUSHRA software etc. to the member of Task 4). A GIT repository is available at: https://github.com/audiolabs/webMUSHRA. However, you have the choice to develop your own MUSHRA test software.

2.4 Listening experiment management

- Conduct a listening experiment with the DSM Course students (Listening experiment participants CLICK) which will be available in July. You have to coordinate with the students via email.
- Make a coordination with the listening experiment room responsible people (This will be done including us).

2.5 Report and Coordination

- The group coordinator will be responsible for the coordination, preparation of the presentation, and final report submission.
- The group coordinator will communicate with the members, collect all results, analyze the results, prepare results between objective and subjective evaluation, and present them as a comparative study and findings of the project.

- Prepare the GIT repository to upload all the audio files, source codes used and developed, results (in the form of tables or graphs), final presentations (power point), etc. The example name of the GIT repository: VC_Seminar_Project_GroupX.
- Collect the Jupyter notebooks (completed by each member for their part), merge together all the notebook information, prepare the final notebook, and upload it as an assignment.
- Other coordination tasks as necessary.

3 Task 1

3.1 Download of audio

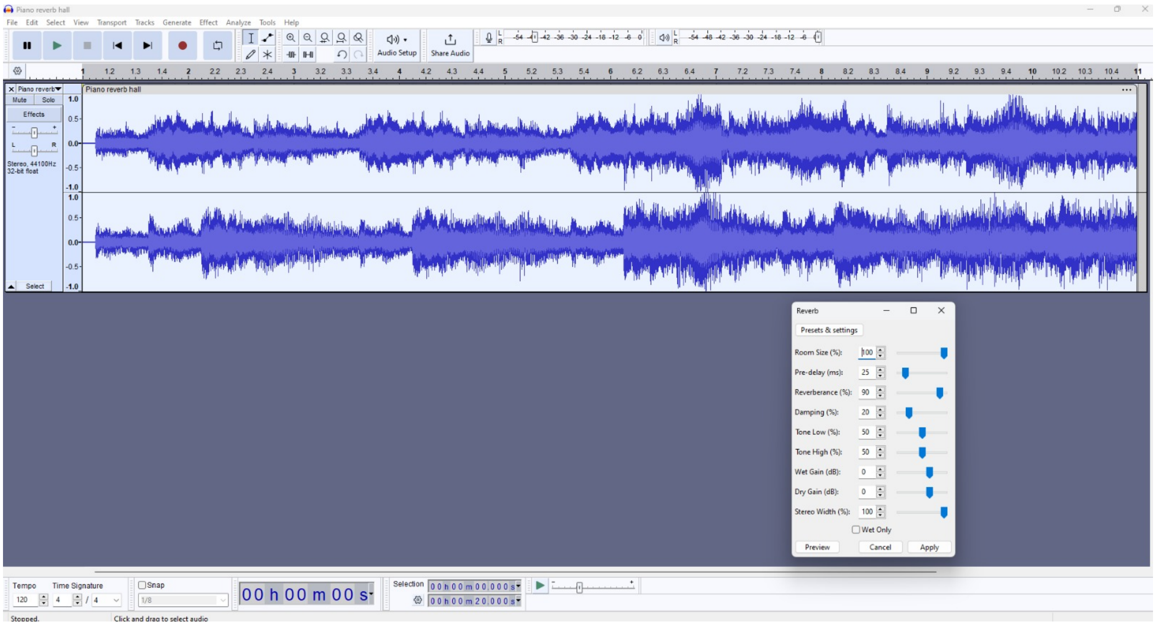


Figure 1: Audio Editing in Audacity

- We obtained five dry audio files with a 24-bit rate, 20 seconds in length, and a sample rate of 44100 Hz from freesounds.org with a Creative Commons license, so they are free to use for study purposes. References to the sounds and licenses can be provided if required.
- The sounds are:
  - Drumbeat
  - Instrumental
  - Ambient walking sound
  - Poetry
  - Piano
- Each sample was then modified in the Audacity software to add reverb.
- In total a set of 20 sounds are present with the following versions for each of the above-mentioned base type: Dry, Opera hall, Reverb hall and Small office room

Below is the idea that was deployed to obtain the above sound samples.

3.2 Audacity settings for synthesizing

To create the effect of reverb in Audacity 1, adjust the reverb settings carefully as follows 2

- **Room Size:** Set this to a high value for large halls and low for small rooms.
- **Pre-Delay:**This simulates the time it takes for the first reflections to reach the listener after the direct sound.
- **Reverberance:**This controls the overall amount of reverb.
- **Damping:** High-frequency cut.
- **Tone Low:** Set this to around 50%. This controls the low-frequency content of the reverb.
- **Tone High:** Set this to around 50%. This controls the high-frequency content of the reverb.
- **Wet Gain:** This controls the volume of the reverberated signal.
- **Dry Gain:** This controls the volume of the original signal.
- **Wet Only:** Ensure this is unchecked unless you want to hear only the reverb effect without the original audio.

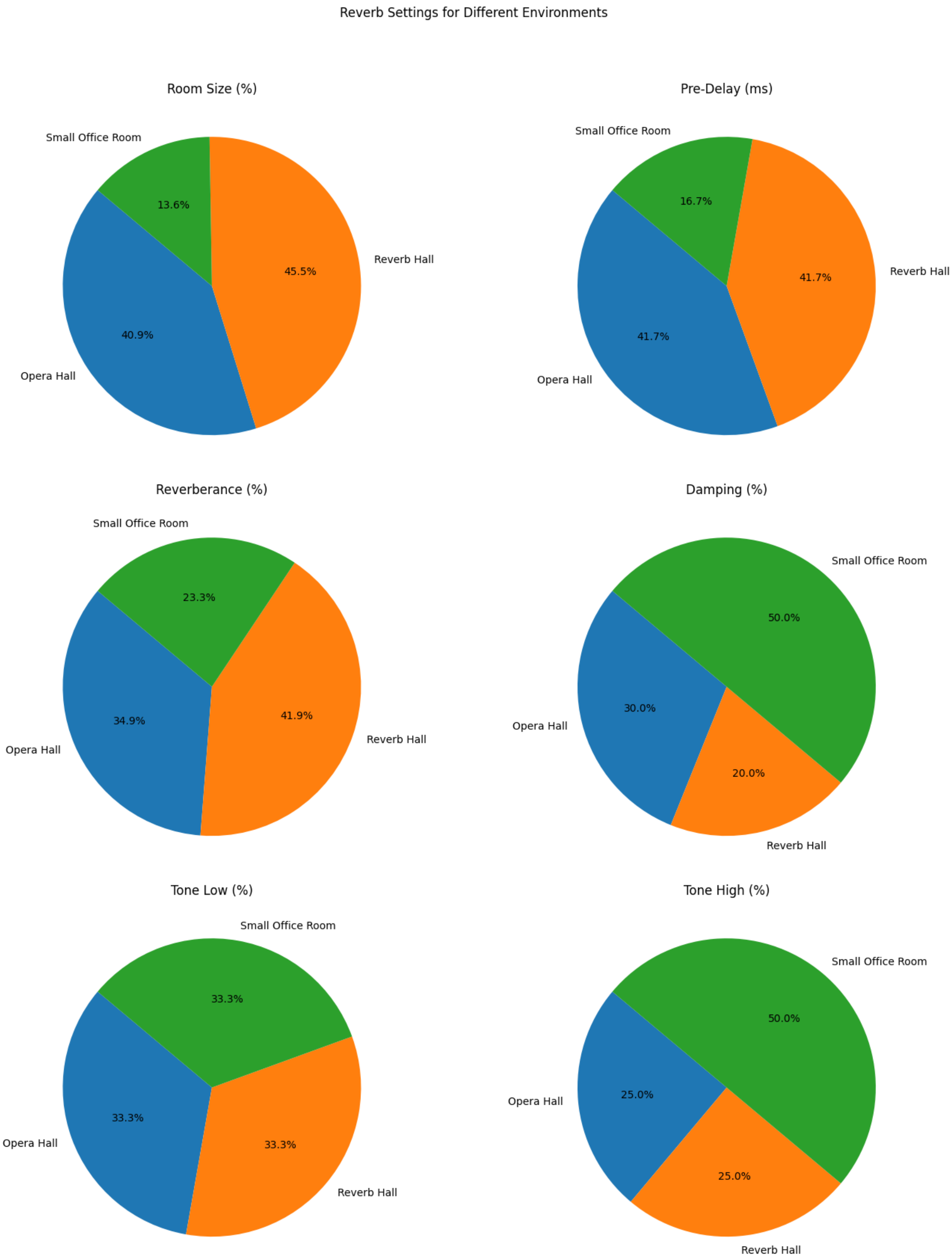


Figure 2: Pie chart - Settings for Opera Hall, Reverb Hall and a small office room

4 Task 2

4.1 Brief overview

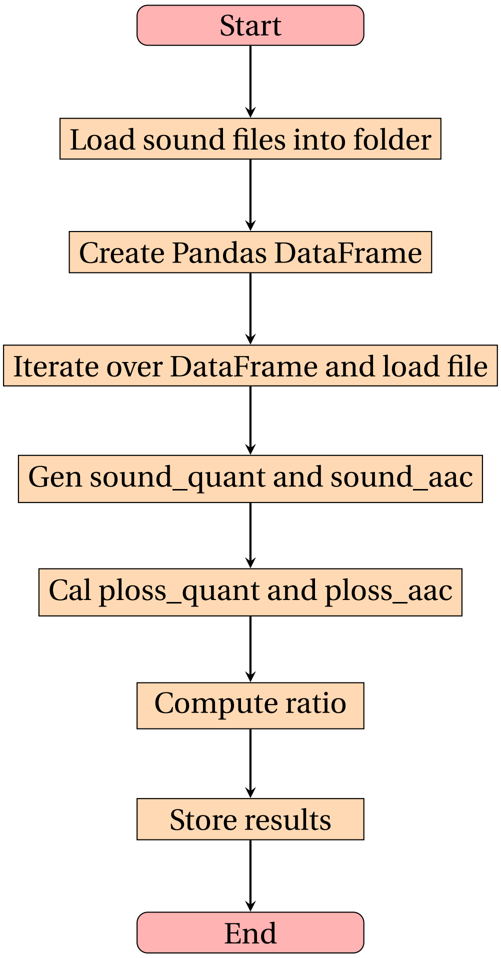
- Collate all sound files into one folder with genre added and display as Pandas DataFrame.
- Iterate over the DataFrame to select a sound and find the `sound_quant` and `sound_aac`.
- Find the perceptual loss (`ploss`) between `sound` and `sound_quant`.
- Find the `ploss` between `sound` and `sound_aac`.
- Compute the ratio of `ploss_aac` with `ploss_quant` and finally store.

Algorithm and Flowchart

4.1 Algorithm and flowchart

1. Load all sound files into a folder and create a Pandas DataFrame with genre information.
2. **For each** sound file in the DataFrame:
   a) Load the sound file.
   b) Generate `sound_quant` and `sound_aac` from the original sound.
   c) Calculate `ploss_quant` between the original sound and `sound_quant`.
   d) Calculate `ploss_aac` between the original sound and `sound_aac`.
   e) Compute the ratio: ratio $= \frac{\text{ploss\_aac}}{\text{ploss\_quant}}$.
   f) Store the results (including the ratio) in a results DataFrame.
3. Save the results DataFrame for further analysis.

Flowchart



4.2 Results

The results are shown as a figure 3 and a table 2 explaining the perceptual loss computed between

- sound and quantised sound
- sound and aac encoded sound

The ratio is also computed to check the quality of compression from the original sound.

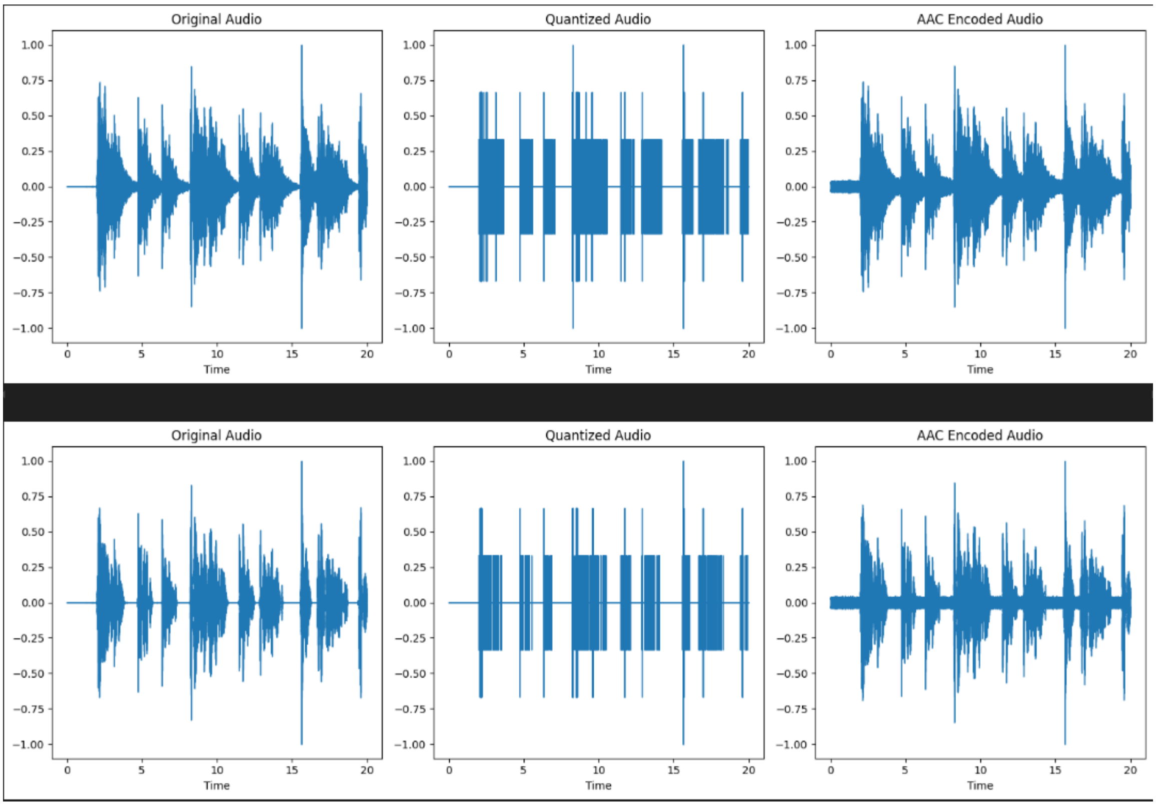


Figure 3: Figure Depicting two samples of Original, Quantised and AAC encoded sound

The table below 2 summarizes the perceptual loss computation results. Each audio file has been processed to generate its quantized version (`sound_Q`) and AAC-encoded version (`sound_A`). The perceptual loss (`ploss`) was calculated for both versions, and the ratio of `ploss_A` to `ploss_Q` was computed.

Additionally, we infer the following from the experiments that were done numerically. The results gave us some fascinating insights as to how humans could perceive audio and how well machines can match the perception computationally. Initial observations note that AAC encoding outperforms the quantization technique comfortably performing best for instrumental and worst for drum beat in relative terms.

We aim to compare the results subjectively (Task 3 and Task 4) and will discuss this further.

1. **Environmental Impact:**
   - Perceptual loss varies across environments (Opera Hall, Reverb Hall, Small Office, Dry), influencing audio quality.
2. **AAC Encoding Influence:**
   - AAC encoding generally increases perceptual loss compared to raw quantization, especially notable in genres like "Poetry"
3. **Genre Sensitivity:**
   - Different genres exhibit varying degrees of sensitivity to environmental and encoding changes, affecting perceived audio fidelity.
4. **Practical Implications:**
   - Understanding these factors is crucial for optimizing audio processing workflows to balance file size reduction with maintaining high audio quality.
5. **Future Directions:**
   - Further research could focus on refining encoding strategies and environmental optimizations to minimize perceptual loss in audio applications.

Table 2: Perceptual Loss Computation Results

	file_name	genre	ploss_Q	ploss_A	ploss_A/Q	winner
1	Poetry opera hall	Poetry	25.49973	49.75037	1.95102	Quantization
2	Poetry dry	Poetry	34.61558	84.35349	2.43686	Quantization
3	Poetry reverb hall	Poetry	68.80215	41.36027	0.60115	AAC
4	Poetry small office	Poetry	21.67172	73.43906	3.38870	Quantization
5	Ambient dry	Ambient	33.58619	3.32906	0.09912	AAC
6	Ambient opera hall	Ambient	30.42014	2.91695	0.09589	AAC
7	Ambient small office	Ambient	31.20757	3.08435	0.09883	AAC
8	Ambient reverb hall	Ambient	32.67670	3.48392	0.10662	AAC
9	Piano reverb hall	Piano	4094.59888	90.99915	0.02222	AAC
10	Piano opera hall	Piano	4001.06494	84.65530	0.02116	AAC
11	Piano small office	Piano	3698.06055	84.45630	0.02284	AAC
12	Piano dry	Piano	3642.72363	83.18191	0.02284	AAC
13	Drum beat dry	Drum_Beat	101.43777	10.88957	0.10735	AAC
14	Drum beat reverb hall	Drum_Beat	23.75473	1.20806	0.05086	AAC
15	Drum beat opera hall	Drum_Beat	26.73443	2.04963	0.07667	AAC
16	Drum beat small office room	Drum_Beat	20.94945	4.16536	0.19883	AAC
17	Instrumental Small office	Instrumental	1153.50464	54.10732	0.04691	AAC
18	Instrumental opera hall	Instrumental	993.65002	48.13313	0.04844	AAC
19	Instrumental Dry	Instrumental	1282.90295	54.42937	0.04243	AAC
20	Instrumental reverb hall	Instrumental	926.78229	44.24751	0.04774	AAC

The above image also summarizes the power of AAC encoding nicely. As evident from this bar plot, one may observe that the Quantization sound was better only on three occasions, whereas AAC-encoded sound fared better for the rest of the sounds.

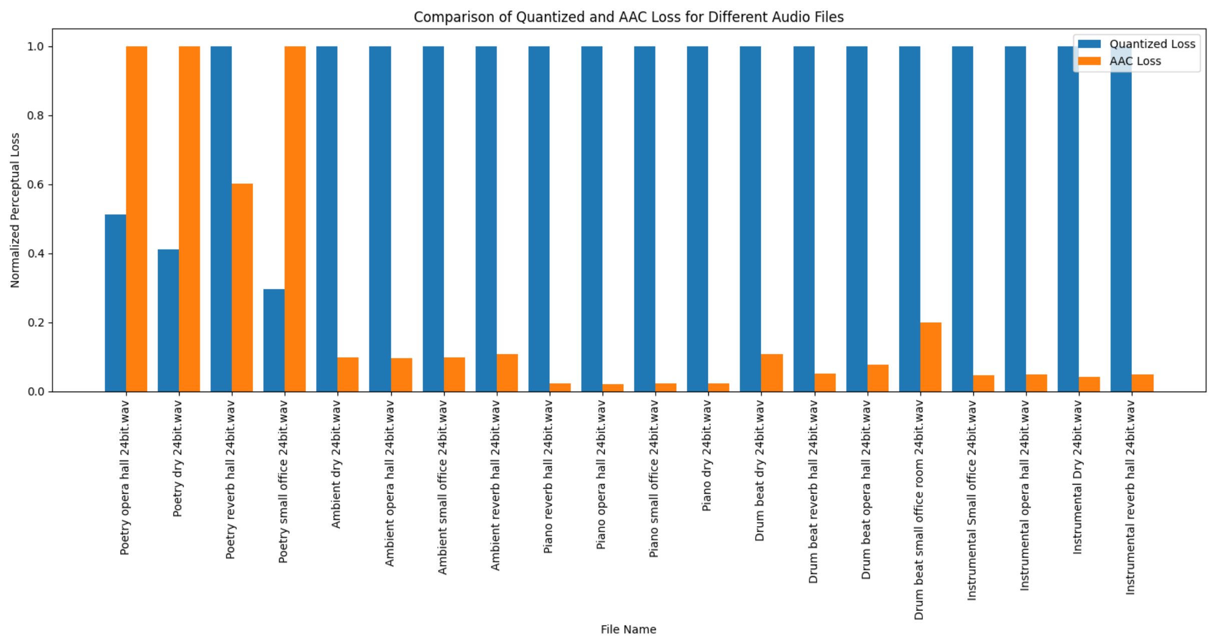


Figure 4: Bar Plot comparing the effectiveness of AAC encoding against Quantized audio

4.3 Discussion and Conclusion

The perceptual loss computation results presented in Table 2 provide valuable insights into the impact of quantization and AAC encoding on different genres of audio files. The primary observation is the distinct behavior of various genres under the two types of processing, leading to different perceptual loss values and, consequently, different winners in terms of the least perceptual loss.

4.3.1 Genre-Specific Observations

- **Poetry:** For poetry files, quantization generally resulted in lower perceptual loss compared to AAC encoding. This can be attributed to the nature of poetry audio, which often involves clear and precise speech with minimal background noise. The high perceptual loss for AAC in this genre suggests that the encoding process might introduce artifacts or distortions that are more noticeable in spoken word audio.
- **Ambient:** In contrast, AAC encoding consistently outperformed quantization for ambient audio files. Ambient sounds often have complex textures and less structured content, which might be better preserved by AAC's advanced compression algorithms compared to simple quantization.

- **Piano:** The piano audio files exhibited extremely high perceptual loss values for quantization, indicating significant degradation of audio quality. AAC encoding, however, maintained much lower perceptual loss, making it the clear winner for this genre. The harmonic and dynamic nature of piano music likely benefits from the sophisticated compression techniques employed by AAC.
- **Drum Beat:** Similar to ambient sounds, drum beat audio files showed lower perceptual loss with AAC encoding. The rhythmic and percussive elements of drum beats might be less susceptible to the types of distortions introduced by AAC.
- **Instrumental:** Instrumental audio files also favored AAC encoding, with significantly lower perceptual loss compared to quantization. This can be attributed to the complexity and variability of instrumental music, which AAC encoding handles more effectively than quantization.

4.3.2 Factors Influencing Perceptual Loss

Several factors contribute to the observed differences in perceptual loss between quantization and AAC encoding:

- **Audio Content:** The type of audio content (e.g., speech, music, ambient sounds) significantly affects the perceptual loss. Content with clear structure and lower complexity (e.g., speech) might be more adversely affected by AAC encoding artifacts.
- **Encoding Algorithms:** AAC encoding uses advanced algorithms designed to preserve perceptual quality by prioritizing audible frequencies and masking inaudible distortions. Quantization, being a simpler process, does not have such sophistication, leading to higher perceptual loss in complex audio.
- **Frequency Components:** Different genres have varying frequency components and dynamic ranges. AAC encoding is better equipped to handle a wide range of frequencies and dynamics, evident in the lower perceptual loss for genres like piano and instrumental music.
- **Compression Artifacts:** The artifacts introduced by AAC encoding can vary depending on the audio content. For example, spoken word (poetry) might reveal compression artifacts more clearly than continuous ambient sounds.
- **Subjective Perception:** Perceptual loss is inherently subjective, and certain types of distortions might be more noticeable in specific genres based on human perception.

4.3.3 Conclusion

The analysis indicates that AAC encoding generally provides lower perceptual loss for most genres except for poetry, where quantization results in better perceptual quality. This suggests that the choice of encoding method should be tailored to the specific type of audio content to ensure optimal audio quality. For genres with complex and rich soundscapes, such as piano and instrumental music, AAC encoding is preferable. In contrast, for spoken word content like poetry, simpler quantization may be more effective in preserving audio quality.

5 Task 3

5.1 Introduction and Objectives

MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) is a popular method for evaluating audio quality, especially for medium to high-quality codecs. It involves listeners rating various versions of an audio sample, including hidden references, on a scale from 0 to 100. Standardized by the ITU under recommendation ITU-R BS.1534, MUSHRA ensures reliable and detailed audio quality assessment.

The objective of this test was to evaluate and compare the perceived audio quality of different sound types processed through various acoustic environments. The focus was on understanding how these environments influence listener perception of sound quality.

5.2 Stimuli

- **Sound Types:**
    1. Ambient
    2. Drum Beat
    3. Instrumental
    4. Piano
    5. Poetry
- **Acoustic Environment:**
    1. Dry/ Original Sound
    2. Piano Hall
    3. Reverb Hall
    4. Small Office
- **Reference and Anchor:** Original high-quality recording of each sound type

5.3 Configuration Steps

1. **Installation of SCALE and MATLAB:**
    - Obtain the latest version of SCALE from the official website.

- Follow the installation instructions for your operating system.

Once both are downloaded open SCALE software while using MATLAB. By setting the current MATLAB folder to the directory where we extracted the program SCALE, we can see a (.m) file there.

2. **Setting Up the Project in SCALE:**
   - Open SCALE and start a new project for the MUSHRA test.
   - Configure Test Parameters:
     - Test Type: Select MUSHRA format
     - Stimuli Type: WAV
     - Bit Depth: Set to 24-bit for all files
   - Import Audio Files
     - Load the reference, anchor, and processed audio samples into SCALE.
     - Organise audio files by sound type and environment.
3. **Designing the Test Interface:**
   - User Interface Layout:
     - Set up sliders for participants to rate each stimulus on a scale from 0 to 100.
     - Ensure a clean, intuitive interface layout.
   - Instructions:
     - Provide clear instructions for participants, including how to use the interface and rating criteria.
4. **Pilot Testing:**
   - Conduct a Pilot Test:
     - Run a pilot test with a small group of participants to ensure setup is functional.
     - Collect feedback to make necessary adjustments.

5.4 Screen Captures

5.4.1 Test Creation

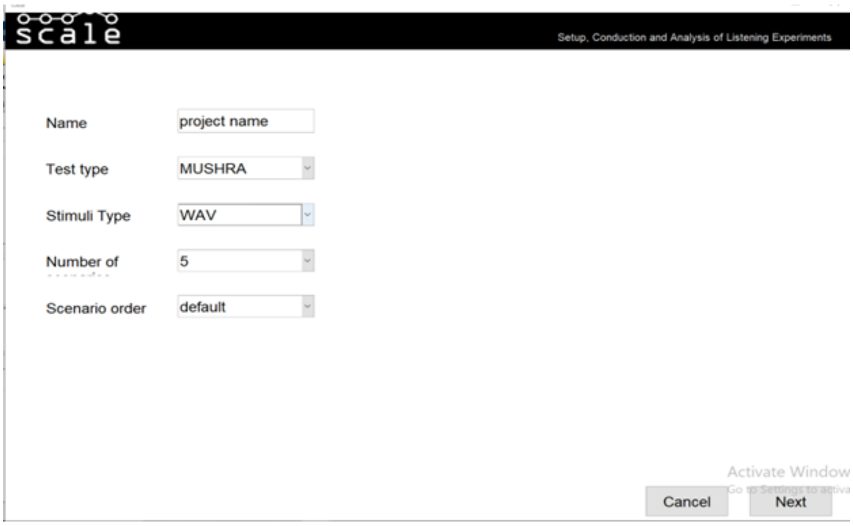


Figure 5: Test Creation 001

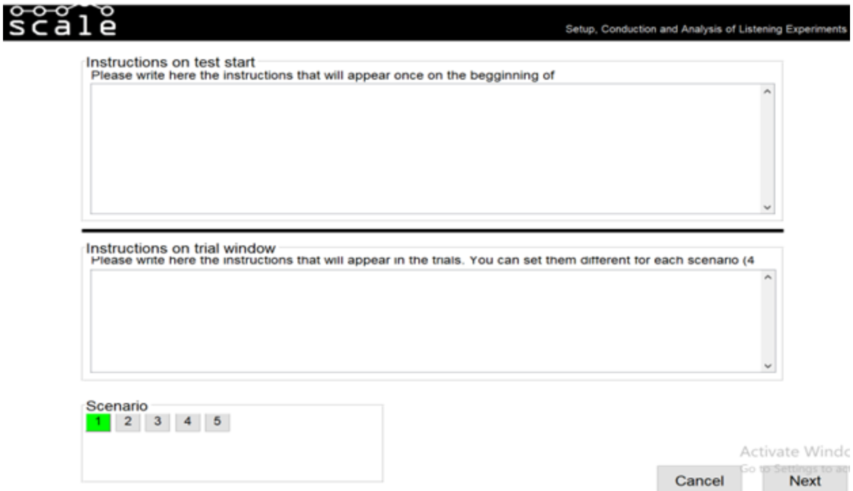


Figure 6: Test Creation 002

5.4.2 Test Performance

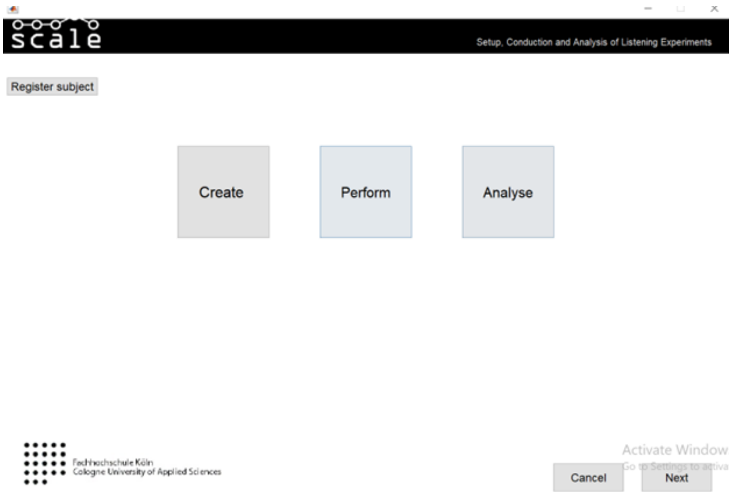


Figure 7: Homepage

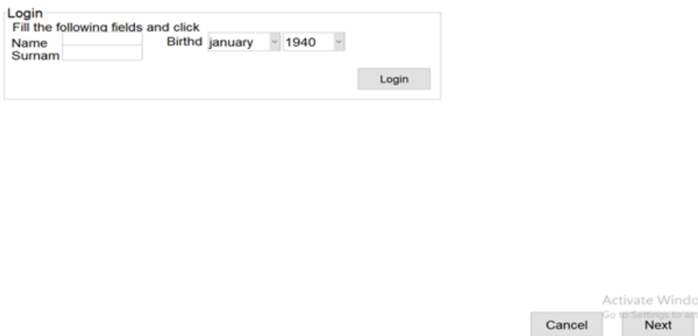


Figure 8: Login Page for Participants

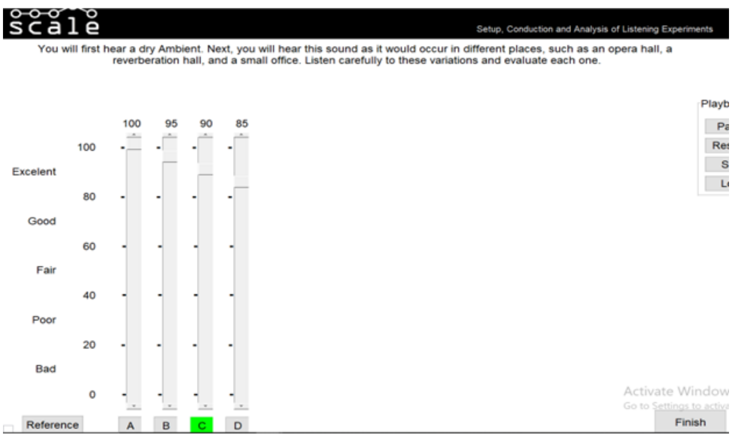


Figure 9: Test Performance / Scoring

6 Task 4

6.1 MUSHRA Listening Test Conduction

The perceptual loss test was conducted in the university media-lab using the MUSHRA methodology, involving 19 subjects to assess audio quality.

The test involved playing a series of audio samples. Subjects were instructed to rate the quality of each sample on a scale, enabling a comparative analysis of perceptual loss across different audio processing methods.

- **Location:** University Media Lab
- **Environment:** Controlled listening environment, Isolated from external noise
- **No. of Participants:** 19
- **Participant Age Group:** 19-38
- **Selection Criteria:** Participants with normal hearing capabilities
- **Equipment Provided in Lab:** High-quality, circum-aural headphones

Each participant was first familiarised with the MUSHRA interface and the nature of the task. Following the training, subjects evaluated the audio samples in randomized order to mitigate any potential bias. The data collected from these ratings was then analyzed to identify perceptual differences in audio quality. The approach during the test was kept structured and organised. Effort was made to control conditions in the lab for the reliability and accuracy of the perceptual loss test results.

The test was conducted over two days. The ratings collected from the participants were stored in a CSV file, and a graph of the individual audio samples was created for further analysis.

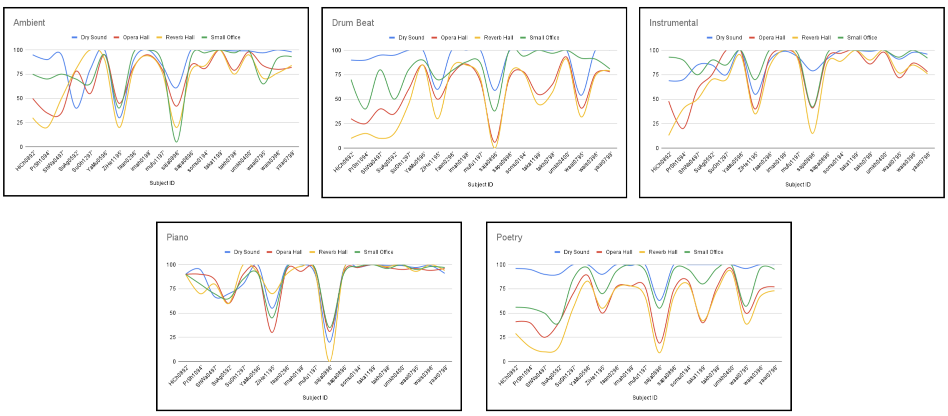


Figure 10: MUSHRA Experiment Results after Outlier Removal

7 Task 5

7.1 Description

This analysis overviews the perceptual differences and similarities in audio stimuli preferences among 19 subjects.

7.2 Indentifying Outliers

To find outliers Interquartile Range (IQR) method was used.

- **Lower Range:** First Quantile-1.5(IQR)
- **Upper Range:** Third Quantile+1.5(IQR)

7.2.1 Detection and Removal

Using IQR analysis, we discovered that two subject IDs showed deviations in specific audio files. Additionally, two other subject IDs deviated the most across all audio files. Consequently, we removed these subjects from the dataset, leaving us with 17 data samples.

7.2.2 Post Outlier Removal Graph

Figure 11: MUSHRA Experiment Results

7.3 Insights and Conclusion

The graphs illustrate that, in some cases(Poetry, Drum Beat), subjects were able to recognise differences between the audio files of various acoustic environments presented to them. This suggests that the alterations made to the Genres were perceptible to the listeners.

However, other cases(Piano, Instrumental and Ambient), all acoustic environments of the audio were perceived as being similar by the subjects. This implies that the changes made to these some Genres did not affect the subjects' perception. Hence, there was minimal perceptual loss, meaning that the integrity and quality of the audio were perceived to be maintained across its different environments.

7.3.1 Acoustic Environment Preferences

- **Dry Sound:** Generally preferred, with fewer outliers, indicating consistent liking.
- **Reverb Hall and Small Office:** These have lower mean scores, suggesting lower overall preference.
- **Opera Hall:** Moderate preference with fewer outliers.