



Audio Engineering Society

Convention Paper

Presented at the 127th Convention
2009 October 9–12 New York, NY, USA

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Subjective evaluation of mp3 compression for different musical genres

Amandine Pras¹, Rachel Zimmerman², Daniel Levitin² and Catherine Guastavino¹

¹ Centre for Interdisciplinary Research in Music Media and Technology,
Multimodal Interaction Laboratory, McGill University, Montréal, Québec, H3A 1EA, Canada
amandine.pras@mcgill.ca, catherine.guastavino@mcgill.ca

² Centre for Interdisciplinary Research in Music Media and Technology,
Department of Psychology, McGill University, Montréal, Québec, H3A 1B1, Canada
rachel.zimmerman@mail.mcgill.ca, daniel.levitin@mcgill.ca

ABSTRACT

Mp3 compression is commonly used to reduce the size of digital music files but introduces a number of potentially audible artifacts, especially at low bitrates. We investigated whether listeners prefer CD quality to mp3 files at various bitrates (96 kb/s to 320 kb/s), and whether this preference is affected by musical genre. Thirteen trained listeners completed an A/B comparison task judging CD quality and compressed files. Listeners significantly preferred CD quality to mp3 files up to 192 kb/s for all musical genres. In addition, we observed a significant effect of expertise (sound engineers vs. musicians) and musical genres (electric v.s acoustic music).

1. INTRODUCTION

The *Motion Picture Experts Group* (MPEG) designed the coder-decoder ("codec") soundfile format mp3 in 1991 as a lossy audio standard to reduce the size of digital music files. In 1991, few people had Internet connections and those who did used dial-up services with typical transmission rates of 5kB/s. Recordable compact discs were not even available in many high-end recording studios. For these reasons, reducing the file size was especially important at that time. Whereas mp3 compression reduces the amount of storage space, the trade-off is a potential decrease in the sonic quality of the audio files. Although these restrictions have eased, many people listen to all their music in compressed formats. In addition, e-tailers encourage people to buy compressed music over the Internet, while the mastering of music productions is still performed at CD quality or even higher quality.

When encoding in mp3, users choose a bitrate (measured in kbits/s) that determines the compression ratio and subsequently the size of the encoded file. During the encoding process, the audio content is first decomposed into 32 frequency sub-bands that are processed based on perceptual models [6]. These models are based on psychoacoustic mechanisms such as the frequency response of human hearing, as well as frequency masking and temporal masking. Then, each sub-band is encoded with a bit allocation determined by the number of available bits (bitrate) and the audio content of the sub-band [1]. This technique theoretically introduces errors in high frequency resolution, as well as noise due to the quantification errors. Regarding audible artifacts introduced by perceptual coders, Brandenburg [1] states that "...the signal may be sounding distorted, but not like harmonic distortion, noisy [...] and rough".

Previous research has compared different types of audio compression algorithms including mp3 with one another [7]. However, perceptual differences between CD quality and mp3 compression have received limited attention. Ruzanski [3] observed a significant interaction between the genre of the clip and how tolerant it was to compression before listener preference would be influenced, for bitrates ranging between 32 and 192 kb/s. Furthermore, Sutherland [8] observed that expert listeners, defined as professional sound engineers with more than ten years of experience, significantly preferred CD quality to compressed files even at very

high bitrates (up to 320 kb/s), whereas Salimpoor [5] found that average listeners were only able to distinguish between very low bitrates (96 kb/s) from CD quality.

Based on these studies, we hypothesize that trained listeners who have studio experience (musicians and young sound engineers with less than ten years of experience) will be able to discriminate between CD quality files and mp3 compressed files and that they will tend more to prefer CD quality files. Furthermore, we seek to know if the results are a function of musical genre and/or listeners' expertise. Finally, we wish to investigate if trained listeners can identify perceptual sound artifacts that are introduced by mp3 compression and if these artifacts depend on the musical genre.

2. METHODS

2.1. Participants

Thirteen trained listeners, eight males and five females, took part in the study and received \$20 for their participation. All subjects passed a standard audiometric for normal hearing. The participants' mean age was 28 (S.D. = 5.6), and mean length of studio experience was six years (S.D. 5.2). They had different sources of studio listening expertise: four as musicians, eight as sound engineers (three young professionals and five students in a sound recording masters degree program), and one as a researcher¹. All participants reported having taken music lessons for 15 years on average (S.D. 4.7) except for one professional sound engineer, who hadn't received musical training. The participants reported listening to music for an average three hours a day (S.D. 2.8) at home or work, and all musical genres were represented. They reported listening to music in lossy compressed format (mp3 or others) 36% of the time on average (S.D. 23.6) as opposed to 64% of the time for uncompressed or lossless compression formats.

2.2. Musical excerpts

We selected five short musical excerpts (one musical phrase ranging from 5 to 11 sec.) in CD quality (44.1 KHz, 16 bit) representative of different musical genres: pop, metal rock, contemporary, orchestra and opera (see detail of the musical excerpts in Table 1).

¹ Two of the co-authors were participants in the study, one professional sound engineer (AP) and one research scientist (CG).

Musical genre	Name of the piece (tune)	Composer	Performers (band/orchestra)	Additional Information
Pop	<i>Irish Green</i>	Bart Moore	Slings & Arrows	Produced by Daniel Levitin
Metal rock	<i>Killing in The Name</i>	Rage Against the Machine	Rage Against the Machine	Produced by Garth Richardson
Contemporary	<i>Diffraction</i>	Yoshihisa Taira	Quatuor Ixtla	Produced by Amandine Pras
Orchestra	<i>Symphonie #5</i>	Gustav Malher	Wiener Philharmoniker directed by Pierre Boulez	Deutsche Grammophon
Opera	<i>Lascia ch'io pianga</i>	George F. Handel	Not listed	Anechoic recording by Angelo Farina, downloaded from www.angelifarina.it

Table 1. Detail of the five musical excerpts used in the study

The pop and opera clips were previously used by Salimpoor [5], and the contemporary clip was used by Sutherland [8]. The five CD quality clips were subjectively matched in loudness by two of the participants in a preliminary listening test (by applying attenuation to the louder clips). We encoded the five excerpts as mp3 with the L.A.M.E. encoder (lame.sourceforge.net) in order to be able to compare our findings with those obtained by Salimpoor [5] and Sutherland [8]. Each excerpt was encoded at five different bitrates (96, 128, 192, 256 and 320 kb/s), resulting in six different versions per excerpt. The experiment consisted of 150 trials corresponding to all possible pairwise combinations of the six different versions of the five musical excerpts.

2.3. Procedure

In each trial, participants were asked to listen to both versions as many times as needed and to choose the version they preferred in a double blind A/B comparison task. Each pair was presented twice in counterbalanced order to nullify order effects. The order of presentation across trials was randomized. The duration of the experiment ranged between 60 and 90 minutes per participant, including a break in the middle of the experiment. The experiment took place in the Critical Listening Laboratory of the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT, Montréal, QC, Canada). This ITU standard room provides high quality controlled listening conditions. Stimuli were presented through a digital analog

converter and monitor controller Grace m906 (Grace Design, Boulder, CO, USA), a Classé CA-5200 stereo amplifier (Classé Audio, Lachine, QC, Canada) and B&W 902D loudspeakers (Bowers & Wilkin, Worthing, West Sussex, England).

2.4. Questionnaire

After the listening test, the participants were asked to fill out a three-part questionnaire. The first part included two open (free response) questions asking how difficult they found the test and how they describe the differences between two different versions on an excerpt. We analyzed these open questions using the constant comparison technique [2]. In the second part, we investigated which sound criteria the listeners used to make their decision. For each excerpt, they were provided with a list of seven sound criteria: High frequency artifacts, Reverberation artifacts, Dynamic range, Stereo image, General distortion, Background noise and Transient artifacts. These criteria were derived from Sutherland's [8] collected on expert listeners who were asked after the listening test to answer an open question about the sound criteria they used to discriminate the files in CD quality over mp3. Participants were also invited to explain, comment or add other useful sound criteria. Furthermore, we asked if they were familiar with the musical genre of the clip. The last part of the questionnaire concerned demographic information, musical training, studio experience and listening habits.

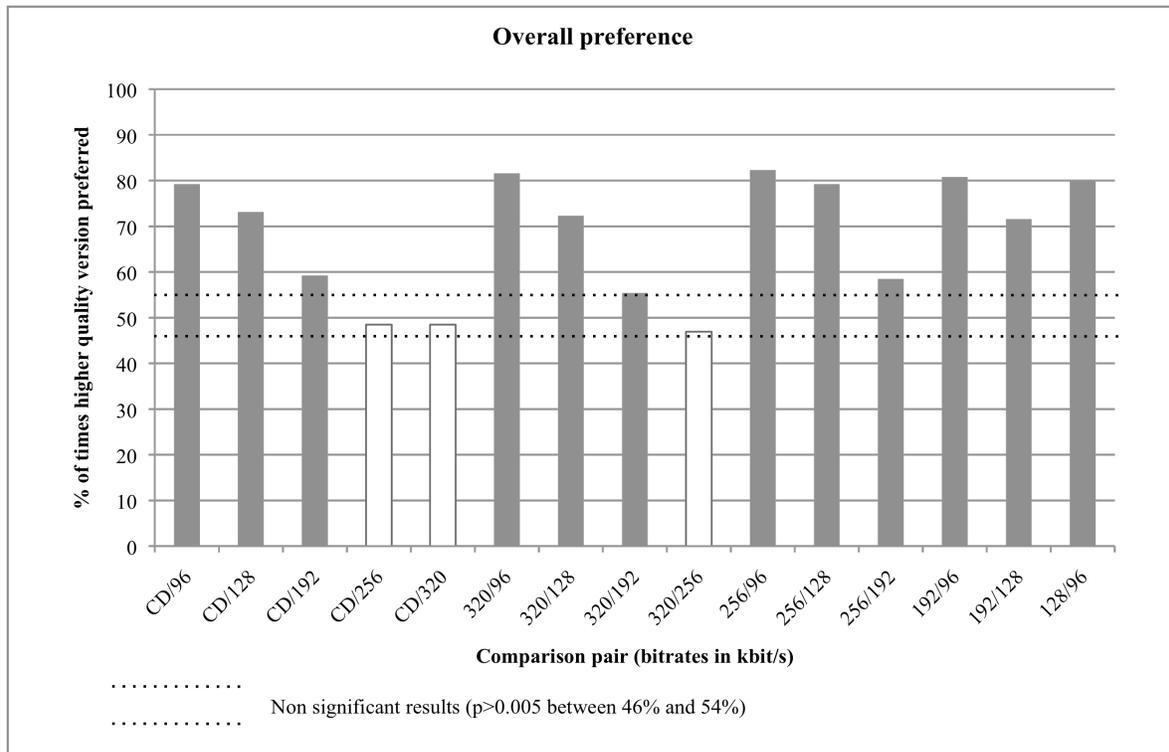


Figure 1. Results collapsed over all participants and musical excerpts, grouped by comparison pair: significant preferences are displayed in grey, non-significant in white.

3. RESULTS

3.1. Overall preference

Over all pairwise comparisons, participants preferred the higher quality version 68% of times. Figure 1 presents the results as percentage of times participants preferred the higher quality version, grouped by comparison pair (results collapsed over all participants and all excerpts). Using the binomial test, performance ranging between 46 and 54% is not significant ($p > 0.05$), meaning that listeners had no significant preference for one version over the other, which strongly suggests that they cannot discriminate between the two. Performance over 54% indicates that listeners could discriminate between the two versions and that they preferred the higher quality. Under 46%, the results indicate that listeners could discriminate between the two versions and they preferred the lower quality version. Over all musical excerpts, listeners significantly preferred ($p < 0.05$) CD quality files to mp3 files for bitrates ranging from 96 to 192 kbits/s. The results are not significant between CD quality files and mp3 files for

higher bitrates (256 and 320 kbits/s). Regarding comparisons amongst mp3 files with different levels of compression, listeners always significantly preferred the higher quality version, except for the comparison between 320 and 256 kbits/s where the results did not reach statistical significance. These overall results demonstrate that mp3 compression does introduce audible artifacts.

3.2. Effect of musical genre and expertise

Our next research question was whether listeners' sensitivity to these audible artifacts is a function of musical genre and listeners' expertise. To test this hypothesis, we compared the distribution of preferences using the chi-square statistic on occurrences (counting the number of times the higher quality version was preferred, and the number of times the lower quality version was preferred). The chi-square test revealed a significant difference between the results' distribution amongst the five musical excerpts ($\chi^2(4) = 22.52$, $p < 0.001$), indicating a highly significant effect of musical genre on preference. However, there was no significant difference between the results' distribution

for pop and metal rock excerpts ($\chi^2(1)=0.06$, $p=0.81$ n.s.) so we combined them under Electric clips. Similarly, there was no significant difference between the results' distribution for contemporary, orchestra and opera excerpts ($\chi^2(2)=4.55$, $p=0.103$ n.s.) so we combined them under Acoustic clips. Finally, a chi-square test confirmed a significant difference between the results' distribution for Electric and Acoustic clips ($\chi^2(1)=17.22$, $p<0.001$), so we present the results for Electric and Acoustic clips separately.

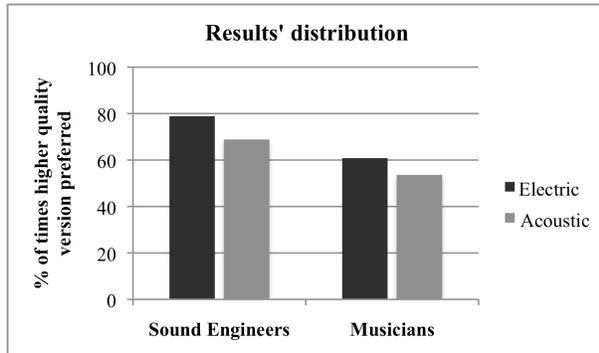


Figure 2. Results' distribution into musical genre and listeners' expertise

A chi-square test revealed a significant difference between the results' distribution amongst the four initial groups of listeners' expertise: musicians, professional sound engineers, sound engineering students and research scientist ($\chi^2(3)=53.69$, $p<0.001$), indicating a highly significant effect of expertise on preference. However, there was no significant difference between professional sound engineers and sound engineering students ($\chi^2(1)=1.29$, $p=0.26$ n.s.). There was no significant difference either between professional sound engineers, sound engineering students and the research scientist ($\chi^2(2)=2.2$, $p=0.33$ n.s.), while there was a significant difference between the musicians and the research scientist ($\chi^2(1)=8.8$, $p<0.01$). Therefore, we combined the results into two expertise levels: one including professional sound engineers, sound engineering students and research scientist, referred to as Sound engineers, and one Musicians group. Here, a significant difference between the Musicians and Sound engineers groups was observed ($\chi^2(1)=44.27$, $p<0.001$), so we present the results separately for each group.

Figure 3 presents the percentage of times Musicians preferred the CD quality files to mp3 files, for every levels of compression (bitrates). Figure 4 presents the percentage of times Sound engineers preferred the CD

quality to mp3, for every levels of compression (bitrates).

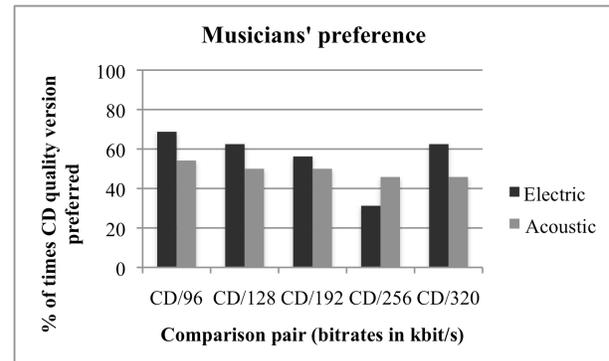


Figure 3. Musicians' preference between CD quality and different levels of compression

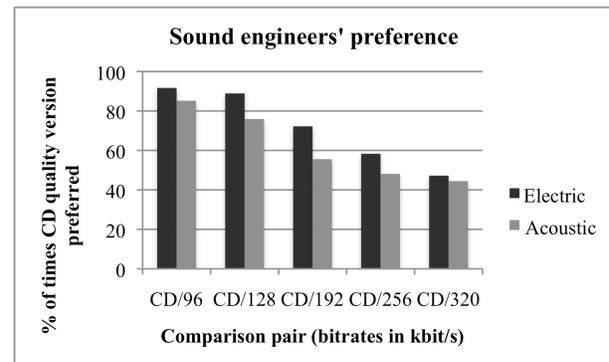


Figure 4. Sound engineer's preference between CD quality and different levels of compression

3.3. Sound criteria

The first part of the questionnaire consisted of two open-ended questions asking participants to describe the differences perceived between the different versions, and the criteria they used to make their judgments. From the free-format responses, we extracted 34 phrasings (mean: 2.6 per participant, S.D. 1.8). These phrasings were assigned to ten non-overlapping categories, seven of which corresponded to the sound criteria proposed in the second part of the questionnaire: High frequency artifacts (6 occurrences), General distortion (5 occ.), Spatial artifacts - that we named Reverberation - (5 occ.), Clarity of the attacks - that we named Transient artifacts - (5 occ.), Stereo image (4 occ.), Dynamic range (2 occ.) and Background noise (1 occ.). Three additional other categories emerged from the free-format responses, namely Liveliness (3 occ.), Articulation of the musical discourse (2 occ.) and Bass artifacts (1 occ.).

The second part of the questionnaire consisted of close-ended questions with free-format justification. For each clip, participants were provided with the list of seven sound criteria derived from previous research and asked whether or not they used each criterion for this given clip, and if so, how and when. Figure 5 shows the distribution of responses expressed in percentage of the total number of sound criteria selected by participants. High frequency artifacts were the most selected criterion, used for all musical excerpts (86%), while Background noise was the least selected (28%). It should be noted that Background noise was not selected for the metal rock clip at all. This can be explained by the fact that the original version (CD quality) of this excerpt is already very noisy (likely because of the guitar amplifiers and effects). Similarly, Dynamic range was rarely selected for the pop clip, likely due to the fact that this musical excerpt sounds extremely dynamically compressed.

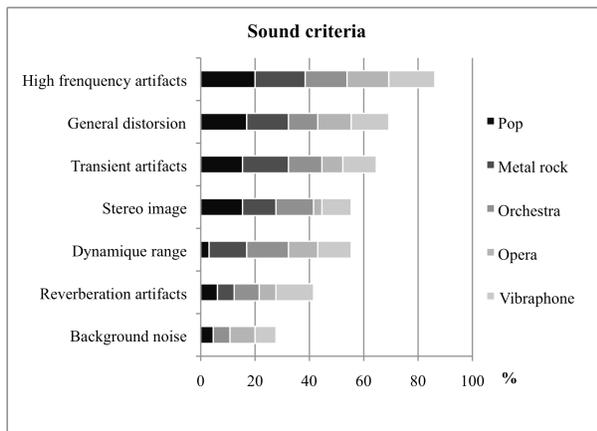


Figure 5. Sound criteria reported by participants to make their decision during the listening test

These results confirm that trained listeners can identify and verbalize the audible artifacts introduced by mp3 compression. However, the distribution of these artifacts does not vary significantly across musical genres ($\chi^2(6)=22.55$, $p=0.92$ n.s. with Yates' correction).

3.4. Difficulty of the listening test

An additional open question asked participants how difficult they found the task. Three reported that was very difficult, one reported that was not too difficult and nine reported that was moderately difficult. Over all musical genres, participants reported that it was hard to focus (1 occ.) and hard to choose between 2 degraded files (1 occ.); the differences between two versions

could be very subtle (4 occ.); two versions of the Acoustic clips (mainly contemporary and opera) were more difficult to discriminate than two versions of the Electric clips (6 occ.). We measured familiarity with the musical genre in the post-questionnaire. Although two participants reported having difficulty with the Acoustic clips because they had never listened to these musical genres, we did not find significant differences between the overall results and the results for familiar genres. However, one musician, a professional drummer, performed significantly better with the pop excerpt (including clear high-hat beats), as opposed to the four others excerpts ($\chi^2(1)=4.13$, $p=0.042$).

4. DISCUSSION

Together, these results indicate that mp3 compression introduces audible artifacts, and that listeners' sensitivity to these artifacts varies as a function of musical genre and listeners' expertise. Specifically, we observed that trained listeners can discriminate and significantly prefer CD quality over mp3 compressed files for bitrates ranging from 96 to 192 kbits/s. Regarding higher bitrates (256 and 320 kbits/s), they could not discriminate CD quality over mp3 while expert listeners, with more years of studio experience, could in the same listening conditions in Sutherland's study [8]. Differences between young sound engineers and experts can be attributed to improved critical listening skills based on individual listening experiences. Furthermore, sound engineers and musicians may not focus on the same sound criteria when listening to music. While sound engineers are trained to hear sound artifacts in general situations, the results from the single case of the professional drummer, who performed significantly better with the excerpt including drums than with the four other excerpts, suggest that musicians are more sensitive to sound fidelity in specific situations. A future study testing professional musicians with excerpts including their personal instrument would allow us to measure audio formats' impact on sound fidelity.

An interesting finding is that the artifacts introduced by mp3 compression were more easily audible on Electric clips (pop and rock, using amplified instruments) than on Acoustic clips (using traditional acoustic instruments). This finding may seem counter-intuitive, as mp3 compression is used predominantly for popular music and less frequently classical music. In the sound engineering community, mp3 format is informally known to require more headroom (difference between

the peak level of the audio signal and the maximum possible level to be quantified) than CD format, although no formal studies on this topic have been identified. Thus, the different results across musical genres could be explained by the use of dynamic compression, often more prevalent in electric music than in acoustic music.

Although listeners' sensitivity to mp3 compression is a function of musical genre, the audible artifacts introduced by mp3 don't depend on musical genre. For all excerpts, High frequency artifacts was reported as the most easily distinguishable artifact, which concurs with theoretical errors in high frequency resolution. Surprisingly, Background noise has been reported as the least distinguishable artifact, even for acoustic clips that have a large dynamic range. Although the reduction of available bits to code the signal intensity is the main technical consequence of audio compression, the noise introduced by the quantification approximation is less perceptible than frequential, temporal and spatial artifacts. Listeners also reported perceptual changes in dynamic range, which can be explained by the reduction of available bits.

Together, these findings suggest that the mastering process of music productions should be adapted to take into consideration the common use of mp3 or other compressed formats. Although the transmission rates of Internet and the storage devices have greatly evolved since the introduction of mp3, the use of compressed formats may continue in early future. This continual need for compressed formats is enforced due to the restrictions of space, time and cost to exchange large amounts of digital information. Furthermore, J. Berger reported an informal study where young listeners preferred compressed formats to CD quality [3]. Therefore, there is a need for mastering processes specifically adapted to lossy compression.

In future studies, we will extend this line of research to quantify the impact of listening conditions on perceptual differences between CD quality and mp3 compression. Furthermore, we will investigate perceptual differences between CD quality and higher quality formats.

5. ACKNOWLEDGEMENTS

The work reported herein was funded in by an FQRSC team grant on the Perception of Audio Quality (P.I.: I. Fujinaga, CG and DJL co-applicants) and a grant from the Grammy Foundation to DJL. Furthermore, the

authors would like to thank Mark Nelson for his comments on earlier drafts, as well as Julien Boissinot, Yves Methot and Harold Kilianski for technical assistance during the experiments conducted at the Centre for Interdisciplinary Research in Music Media and Technology (Montréal, QC, Canada).

6. REFERENCES

- [1] Brandenburg, K. (1999). *MP3 and AAC explained*. Paper presented at the 17th International Conference: High-Quality Audio Coding.
- [2] Glaser, B. G. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research* (p. 271). Chicago: Aldine Pub. Co.
- [3] Jamieson, A. (2009). iPod generation prefers tinny music to CD. Retrieved July, 31, 2009, from <http://www.telegraph.co.uk/technology/apple/4941506/iPod-generation-prefers-tinny-music-to-CD.html>.
- [4] Ruzanski, E. P. (2006). Effects of MP3 encoding on the sounds of music. *Institute of Electrical and Electronics Engineers*, 25(2), 43-45.
- [5] Salimpoor, V. (2006). *Subjective evaluation of popular audio compression formats*. Unpublished manuscript. McGill University.
- [6] Shlien, S. (1994). Guide to MPEG-1 audio standard. *Institute of Electrical and Electronics Engineers*, 40(4), 206-218.
- [7] Soulodre, G., Grusec, T., Lavoie, M., & Thibault, L. (1998). Subjective evaluation of state-of-the-art two-channel audio codecs. *Journal of the Audio Engineering Society*, 46(3), 164-177.
- [8] Sutherland, M. E. (2007). *The evaluation of different compression formats: From amateur to professional listeners*. Unpublished manuscript. McGill University.