# SOUND QUALITY ENHANCES THE MUSIC LISTENING EXPERIENCE

*Victoria J. Williamson* [1], *Michael South* [2] *and Daniel Müllensiefen* [2]

[1] University of Sheffield (UK) [2] Goldsmiths, University of London (UK)

## ABSTRACT

Most people are unable to discriminate the sound quality of music recordings across a range of typical commercial products. However, the effects of very high digital resolution levels on the music listener remain untested. The present study aimed to determine whether people respond differently to recorded music at three resolution levels: 320kbps (mp3), 16-bit/44.1kHz (CD), and 24-bit/192kHz (Studio Master).

In Study 1, participants rated the sound quality of continuous 30-second song segments, presented randomly across the three resolution levels. We replicated the well-established failure to discriminate mp3 and CD, but noted a preference for Studio Master recording. In Study 2, participants scored the sound quality of two songs at the lowest (mp3) and highest (Studio Master) digital resolution levels. We also assessed felt emotion and analyzed listening time as an implicit measure of preference. In all cases participants successfully discriminated between the two resolution levels and preferred Studio Master recording. Implications for future tests of sound quality in music recordings and music listening practices are discussed.

## 1. INTRODUCTION

The history of music recording has been marked by the evolution from analogue to digital formats. Originally, music recordings were made by processes that transformed air molecule vibrations into analogous ('analogue') physical alterations in the reception medium, such as vinyl LP grooves. Modern digital recordings, by comparison, convert the pattern of source vibrations into sequences of 1s and 0s. Modern digital sound recording formats incorporate 'lossy' psycho-acoustic algorithms that allow music file size to be compressed, thereby accommodating portable, low cost music listening technology.

Digital recording of music and subsequent compression is a complex process. Two parameters that determine the resolution of auditory information and that are relevant to the present paper are sampling rate and bit depth. Sampling rate is the number of times per second that the original continuous signal is sampled. For example, CD music is encoded at 44.1kHz, meaning that 44,100 samples per second are extracted from the original signal. Bit-depth determines the range of different values possible when measuring the amplitude of the analogue waveform. For example, the 16-bit depth of CD recordings offers 65,536 discrete values of amplitude. Together, the sampling rate and bit depth describe the digital resolution or 'sound quality' of recording, which for CD is 1411.2 kilobits per second (kbps; sampling rate x bit depth).

In today's music market the most consumption is via highly compressed digital formats, namely mp3 (Krause, North & Hewitt, 2013). The highest sound quality level supported by mp3 is 320kbps, which equates to a compression ratio of around 4:1.

The benefits to the listener of flexible access to thousands of pieces of music in a miniature device are beyond doubt. However, assuming music is played on devices that are capable of reproducing high digital resolution the question remains; does higher sound quality (digital resolution) lead to a richer experience for the music listener?

Research studies have taken one main approach to tackling this question: testing whether people can discriminate between mp3 and CD levels of resolution. Results in this area are consistent and suggest that people are insensitive to the difference. Pras, Zimmerman, Levitin and Guastavino (2009) evaluated CD and mp3 discrimination using pair-wise comparison of 5 -11 second samples of pop, rock, orchestral, contemporary and opera music. Thirteen 'expert' participants (musicians, studio engineers) could not discriminate between CD and mp3 (256kbps and 320kbps levels). Similar negative results were obtained by Pras and Guastavino (2010) and Yoshikawa et al. (1995), although it should be noted that Olive (2011) found that 70% of their student participants and 86% of 'expert' listeners preferred CD to the most highly compressed mp3 (128kbs) sound.

In most cases it seems that standard CD and mp3 formats are difficult to discriminate, even for even highly trained listeners. However, there is suggestive evidence from Olive (2011) that larger differences in digital resolutions yield above chance levels of detection and a preference for higher resolution.

Developments in digital audio technology mean that it is now necessary to extend the standard comparisons of sound quality in music recording to include Studio Master along with traditional mp3 and CD resolution. Studio Master music recordings are typically encoded with a bit-depth of 24-bit and very high sample rates, such as 192kHz (i.e. 4608 kbps), as is the case of the stimuli of the present study.

Psycho-acoustic studies provide evidence that people may be able to discriminate between CD (1411.2 kbps) and Studio Master (4608 kbps) digital resolution thanks to the higher sampling rate. Woszczyk (2003) argued that the temporal resolution for CD (sampling interval of 22.7µs) is insufficient, given that many transient onsets possess a rise time of less than 10µs. By comparison, the 5.2µs sampling interval of Studio Master recording resolves more waveform detail. The human ear could thus discriminate between these two resolution levels if it were able to resolve intervals of c.10µs. Krumbholz, Patterson, Nobbe and Fastl (2003) used backward masking to identify temporal resolution thresholds in the human ear of between 10µs and 20µs.

More recent findings point to another possible mechanism of higher sound quality discrimination: the presence of inaudible high frequencies. Kuribayashi, Yamamoto, and Nittono (2014) tested double blinded discrimination of two short musical excerpts (200 seconds of J. S. Bach's French Suite No. 5) with or without inaudible high-frequency components (above 20 kHz). Participants were unable to identify the difference in these two short samples however, high-α EEG power (10.5-13 Hz) was larger for the last 50 seconds of the excerpt with the high-frequency components than for the excerpt without them. Therefore, inaudible high frequencies, of the kind present in higher sound quality recordings, have an impact on brain activity in the absence of conscious awareness.

These combined findings suggest that the higher sampling rates of Studio Master recording levels might result in differences in physiological response to and/or perceptual experience of music compared to lower levels (mp3 and CD) tested to date.

## 1.1 The present study

The present study built on previous evidence and investigated discrimination of music at three digital resolution levels: mp3 (320kbps), CD (1411.2 kbps) and Studio Master (4608 kbps).

Firstly, we expected to replicate poor listener discrimination between mp3 and CD levels, but we hypothesized that the highest sound quality resolution level (Studio Master) would be detectable compared to the lowest level (mp3).

Secondly, the present paper went beyond measuring sound discrimination since the 'listening experience' is not limited to this dimension. The importance of emotional reactions in music listening is paramount since '*The most common goal of musical experiences is to release emotions*' (Juslin & Vastfjall, 2008, p.559). To date however, we found no research into how sound quality impacts on the emotions felt by the music listener. In order to measure felt emotion in the present study we employed the Geneva Emotional Music Scale (GEMS-9) (Zentner et al., 2008).

Finally, it is important to note that a music listener's experience extends beyond conscious evaluations related to either discrimination or felt emotion. Implicit responses have been recorded successfully in consumer behavior studies, where participants react to music without being consciously aware of its concomitant properties or, in some cases, even its presence (North & Hargreaves, 2008).

The implicit measure of the listening experience considered in the present study was self-determined listening time, a proxy measure of preference. Lindsen, Moonga, Shimojo and Bhattacharya (2011) investigated sampling behaviours and decision-making ("like" or "dislike") for closely matched musical excerpts. In this study people were free to sample excerpts until reaching a decision on their preference. Participants listened to their preferred option ("like") for longer when making their decision in 73% of trials in contrast to a limited sample bias for the "dislike" condition. People were unaware that their listening time was being measured, thereby providing an implicit objective measure of listening preference.

In summary, the present paper compared the impact of three sound quality levels (mp3, CD and Studio Master resolution) on the 'listening experience' as defined by explicit discrimination, levels of felt emotion, and time spent listening. Study 1 comprised explicit rating of continuously changing sound quality in a single song across all three levels (mp3, CD and Studio Master). Study 2 comprised explicit rating of the sound quality of songs reproduced at the two extreme digital resolution levels (mp3 and Studio Master) combined with judgments of felt emotion and an implicit measure of listening preference (listening time).

# 2. METHOD

## 2.1 Materials

Both studies in the present paper took place in a sound attenuating booth. Inner and outer chambers included a 102mm thick acoustic modular panel, separated by an air gap of 100mm. Tested in accordance to ASTM standards, the booth provided minimum noise reduction of 37db and a sound absorption coefficient of 0.38 for the 63Hz octave band, rising to 93db noise reduction and a sound absorption coefficient of 1.06 for the 1kHz octave band, and >93db noise reduction and >1 sound absorption coefficient for >1kHz octave bands.

The music reproduction equipment consisted of a Toshiba laptop providing the digital stream, which was fed via an Ethernet router to a Linn Majik DSM streamer (integrated amplifier), connected via Linn interconnects to Linn Majik 109 loudspeakers. Volume was kept at a constant level, assessed in pilot testing to be a comfortable average listening volume for someone of normal hearing. The system employed the Linn Kinsky/UPnP control systems allowing interface to a data logger for recording of listening time (Study 2).

Four music recordings were used across the two studies. All were sung by Carol Kidd accompanied by guitarist Nigel Clark ("Stormy Weather", "I Got Lost in His Arms", "The Shadow of Your Smile" and "Moon River"). All songs were originally recorded at Studio Master level. The recordings were then prepared to different levels of digital resolution using Magix Sequoia software. To create the CD and mp3 files, the original Studio Master recordings were down-sampled to the appropriate standard sample rates and bit-depths.

For Study 1, "Moon River" was prepared as a looped version (song repeated once, duration of 7 min 48s) consisting of 30-second segments. Each 30s segment was followed by an unnoticeable shift in resolution to the subsequent segment with a different and randomly selected digital resolution; all possible combinations of changes in resolution levels were included. The final 'track' was rendered and exported at the maximum quality of 192kHz/24-bit. The lower quality samples within the track (CD and mp3) were still of low quality after the rendering treatment and high quality export, since digital information had been lost in the down-sampling. The resultant single 'track' presented a seamless repeated song to the listener.

In a break between the two studies participants heard "The Shadow of Your Smile" (Duration: 4 min 49 s) in mp3 format. Like the other habituation phases of the procedure the use of mp3 music in this way helped to establish a constant baseline level of recent music exposure for all participants. For Study 2, "Stormy Weather" (Duration: 3 min 47 s) and "I Got Lost in His Arms" (Duration: 3 min 37 s) were prepared at mp3 and Studio Master levels using the above techniques.

The GEMS-9 was used to assess felt emotion in Study 2. The GEMS-9 presents nine factors with emotion adjectives. These include Wonder (Dazzled, Moved), Tenderness (Affectionate, In love), and Nostalgia (Dreamy, Melancholic). Participants rate each of the nine factors on a 5-point scale from felt "Not at all" to felt "Very much" (Zentner et al., 2008).

## 2.2 Participants

In total 20 participants (8 female) took part in the studies. Participants were graduate students aged between 21 and 59 with a mean age of 27.8 years (SD = 8.5). No reward was offered for participation.

## 2.3 Procedure

Participants were seated centrally in the booth, equidistant from each loudspeaker. We began with a habituation phase where participants completed a musical experiences questionnaire (not reported in this paper) while the mp3 version of "Moon River" played in the background. Participants were instructed to pay no attention to the music but to concentrate upon the questionnaire. The aim of the habituation phase was to establish a baseline for the lowest level of sound quality, in case participants had been listening to music recorded at different levels before the study began.

Participants were then given the Study 1 score sheet, and instructed to listen carefully to the repeated version of "Moon River". They were informed that each version of the song, although seamless, consisted of seven 30s segments at differing 'sound quality' (three mp3, two CD and two Studio Master). They were requested to rate each 30s segment for sound quality on a 1-7 scale. Immediately at the point of each 30s change, visible on the playback device, the participant scored the previous segment for sound quality. They were advised not to score the first segment at 1 or 7 in order to avoid floor/ceiling effects, and to score based on their idea of sound quality, not on performance or enjoyment at any particular point. We pointed out that differences in sound quality would be subtle but the importance of making a decision as soon as possible.

Study 2 began following two minutes of relaxation. First participants completed another habituation phase during which they filled in a personality questionnaires (not reported in this paper) while "The Shadow of Your Smile" played at mp3 level in the background. As in Study 1, the habituation phase served to establish an equal baseline of low sound quality for all participants.

Participants were then given the Study 2 sound quality score sheet and told that they would hear two different songs. Their first task was to score each whole song for level of sound quality on the same 1-7 scale as in Study 1. Each participant was played one of the four order-counterbalanced combinations of two songs ("Stormy Weather" and "I Got Lost in His Arms") at both levels of digital resolution (mp3 and Studio Master).

Participants were informed that they would listen to each song once. After that they were given a remote control and could replay/fast forward/rewind the song as much as they liked in order to make their decision on sound quality. In addition, participants were asked to consider the emotional effect of each song and completed the GEMS-9 for felt emotions. They were asked to rate only the emotions they felt while listening to the music, not emotions conveyed by the music. Again, they were told that they could listen as long as they liked in order to complete the GEMS-9. Total listening time for each song was recorded covertly on the laptop.

# 3. RESULTS

## 3.1 Study 1

For the song Moon River (played twice), each participant provided four ratings for CD conditions, four for Studio Master, and six for mp3, generating a dataset of 280 total observations. The means and standard deviations for the three conditions are given in Table 1.

| Sound Quality Level | Mean Rating | SD |
|---|---|---|
| mp3 | 4.30 | 1.03 |
| CD | 4.24 | 1.09 |
| Studio Master | 4.63 | 1.13 |

**Table 1.** Means and standard deviations for sound quality ratings across the three levels of digital resolution. Ratings were given on a 1-7 scale, with 7 indicating the highest quality level.

In order to determine whether the raw scores represented in Table 1 were statistically significant, we applied linear mixed models with random effects that account for heterogeneous variances across the groups (R package 'nlme').

For fixed effects we included as independent variables; i) the track position of the sound quality condition, as compared to other segments of the same quality ('Position'), ii) whether it was the first or the second play of the track ('Play'), and iii) sound quality level of the 30s segment (mp3, CD, Studio Master). Participants were included as a random effect and variances for the three sound quality conditions were estimated separately to account for the difference in sample size and variance. Subjective ratings of sound quality were the dependent variable.

The results can be seen in Table 2. The difference between the Studio Master and CD conditions was significant ($t\ (255) = 2.67$, $p = .008$). When mp3 was used as the reference level, there was a significant difference between Studio Master and mp3 ($t\ (255) = 2.32$, $p = .021$: not shown in Table 2). None of the remaining effects reached the common significance level.

| Variable | Parameter Estimate | 95% Confidence Intervals | | t-value | p-value |
|---|---|---|---|---|---|
| | | Lower | Upper | | |
| **Sound Quality** | | | | | |
| mp3 | 0.05 | -0.23 | 0.33 | 0.35 | .729 |
| Studio Master | 0.40 | 0.10 | 0.70 | 2.67 | .008 |
| **Position** | | | | | |
| Two | -0.05 | -0.29 | 0.19 | -0.40 | .689 |
| Three | 0.01 | -0.36 | 0.38 | 0.07 | .944 |
| **Play** | | | | | |
| Second | 0.19 | -0.03 | 0.41 | 1.72 | .087 |

**Table 2.** Model parameters and significance levels for mixed effects model of explicit sound quality ratings. In this analysis the model reference levels were CD (Sound Quality), First Position, and the first play of each track.

## 3.2 Study 2

The results of Study 1 indicated that explicit ratings of sound quality were significantly higher for Studio Master compared to CD and mp3, while the ratings of the latter two did not differ from each other. The first aim of Study 2 was to replicate this main effect, using different experimental stimuli (different songs) and presentation paradigm (whole songs rather than 30s excerpts).

The second aim was to assess whether participants' felt emotion ratings and self-determined listening time would differ between sound quality conditions. The means and standard deviations for all three dependent variables are given in Table 3.

| Sound Quality Level | Explicit Rating | Listening Time | GEMS Score |
|---|---|---|---|
| mp3 | 4.35 (1.14) | 179.65 (57.53) | 18.2 (4.07) |
| Studio Master | 5.60 (0.99) | 211.40 (72.71) | 20.7 (4.78) |

**Table 3.** Means (standard deviations) for explicit sound quality ratings, listening time, and GEMS score across the two levels of sound quality.

According to the raw mean figures, participants rated sound quality as higher, felt more emotions, and listened longer to tracks in the Studio Master compared to the mp3 condition. These differences were subsequently tested for statistical significance. To analyze sound quality and emotion ratings we employed two linear mixed-effect models. The two models used sound quality as the main independent variable but also accounted for fixed effects from presentation slot (first vs. second) and song ("Stormy Weather" vs "I Got Lost") while including participants as a random factor.

As can be seen in Table 4, sound quality ratings for Studio Master were significantly higher than those given to mp3 ($p = .001$), while there was no significant difference between the tracks presented first or second nor between the two songs (though this effect bordered on significance; $p = .054$). The mixed model for emotions (Table 5) found an advantage for Studio Master compared to mp3 ($p = .008$), this being the only significant effect in the model.

| Variable | Parameter Estimate | 95% Confidence Intervals | | t-value | p-value |
|---|---|---|---|---|---|
| | | Lower | Upper | | |
| **Sound Quality** | | | | | |
| Studio Master | 1.17 | 0.52 | 1.82 | 3.81 | .001 |
| **Play** | | | | | |
| Second | -0.13 | -0.79 | 0.51 | -0.43 | .670 |
| **Song** | | | | | |
| Stormy Weather | -0.63 | -1.27 | 0.01 | -2.06 | .054 |

**Table 4.** Model parameters and significance levels for the two mixed effects models, for explicit sound quality ratings. In this analysis the reference levels were mp3 (Sound Quality), First Play, and the song "I Got Lost".

| Variable | Parameter Estimate | 95% Confidence Intervals | | t-value | p-value |
|---|---|---|---|---|---|
| | | Lower | Upper | | |
| **Sound Quality** | | | | | |
| Studio Master | 2.67 | 0.79 | 4.55 | 3.00 | .008 |
| **Play** | | | | | |
| Second | -0.68 | -2.55 | 1.19 | -0.77 | .452 |
| **Song** | | | | | |
| Stormy Weather | -1.08 | -2.95 | 0.79 | -1.22 | .239 |

**Table 5.** Model parameters and significance levels for the two mixed effects models, for GEMS emotion scores. Reference levels match those of Table 4.

There was an issue with the analysis of listening time. Many participants opted to listen only up to the end of the song despite the instruction to listen for as long as they liked beyond then. Therefore, listening times were skewed towards 217 and 227 seconds (the play times for the two songs). Because of this skewed data distribution it was not possible to compute a linear model for listening time. Instead, we recoded listening times into three categories: early termination / end of song / beyond end of song and computed a chi square test between sound quality and listening time, using participants as a blocking factor. The association between categorized listening time and sound quality was not significant (chi square = 1.897, $p = 0.138$).

# 4. DISCUSSION

For many years the accepted wisdom with regards to digitally recorded music has been that very few people can tell the difference between the standard commercially available sound resolution levels, namely CD and mp3. The results from Study 1 support this assertion: ratings of sound quality across CD and mp3 resolutions did not differ. This finding is also in line with previous literature on the subject (Yoshikawa et al., 1995; Pras et al., 2009; Pras & Guostavino, 2010).

However, the development of music reproduction technology and a trend towards commercial availability of high resolution recordings, such as Studio Master, has stimulated new research questions – can people detect a difference in sound quality between lower (mp3) and much higher digital resolution (Studio Master) music recordings? Moreover, does such an increase in resolution impact on the music listening experience in terms of felt emotions and/or listening preferences.

Participants in the present study consistently rated Studio Master music as higher in subjective sound quality compared to mp3, both in terms of 30s excerpts in a continuous song (Study 1) and across complete songs (Study 2). Furthermore, participants reported experiencing more felt emotions to Studio Master level music compared to the same songs heard at mp3 level.

Our covert measure of self-determined listening time did not reach significance in the present modelling due to the fact that most participants chose to listen to the complete song rather than spontaneously extending their listening time. However, in terms of the raw numbers this implicit measure of listening preference (Lindsen et al. 2011) also indicated a preference for Studio Master.

This early research into the impact of sound quality on the music listening experience should be treated with caution. At present we have no evidence regarding the source of the differences found between Studio Master and lower levels of digital resolution. Previous literature suggests possible mechanisms of effect that include, but are not limited to, a higher sampling rate (Woszczyk, 2003) and/or the presence of inaudible higher frequencies (Kuribayashi, Yamamoto, & Nittono, 2014) in higher quality Studio Master recordings. These mechanisms should now form the basis for future controlled experiments that explore *how* enhanced sound quality might impact on the music listening experience.

We hope that this brief report stimulates future studies into the impact of high levels of musical sound quality on the listening experience that go beyond both traditional comparison stimuli (mp3 and CD) and methods (short excerpts and objective ratings). Such research would have implications not only for private music consumption but also for the ubiquitous public spaces which rely on positive explicit and implicit reactions to musical sound (North & Hargreaves, 2008).

## 5. ACKNOWELDGEMENTS

This research was possible thanks to an industry partnership between Goldsmiths, University of London and Linn Records. We thank Robert Cammidge, who created the music files, and Neil Parry, who installed the equipment on which the stimuli were presented. We also thank Caroline Dooley and Gilad Tiefenbrun for their support in organizing the logistics of the project. We declare that Linn Records had no influence on the design of the study, the analysis of the results or the interpretation of the data.

## 6. REFERENCES

Juslin, P.N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31, 559 - 575.

Kuribayashi, R., Yamamoto, R., & Nittono, H. (2014). High-resolution music with inaudible high-frequency components produces a lagged effect on human electroencephalographic activities. *Neuroreport*, Apr 9. [Epub ahead of print]

Krause, A.E., North, A.C., & Hewitt, L.Y. (2013) Music-listening in everyday life: Devices and choice. *Psychology of Music*, Published online August 14, 2013, doi: 10.1177/0305735613496860

Krumbholz, K., Patterson, R.D., Nobbe, A., & Fastl, H. (2003). Microsecond temporal resolution in monaural hearing without spectral cues? *Journal of the Acoustical Society of America*, 113(5), 2790-800.

Lindsen, J.P., Moonga, G., Shimojo, S., & Bhattacharya, J. (2011). Swayed by the music: Sampling bias towards musical preference distinguishes like from dislike conditions. *Consciousness and Cognition*, 20, 1781 – 1786.

North, A & Hargreaves, D. (2008) *The Social and Applied Psychology of Music*. Oxford University Press: USA.

Olive, S. (2011). Some new evidence that teenagers may prefer accurate sound reproduction. *Audio Engineering Society Convention: 131*, Paper Number: 8583.

Pras, A., & Guastavino, C. (2010). Sampling rate discrimination: 44.1 kHz vs. 88.2 kHz. *Audio Engineering Society Convention: 128*, Paper Number: 8101.

Pras, A., Zimmerman, R., Levitin, D., & Guastavino, C. (2009). Subjective evaluation of mp3 compression for different musical genres. *Audio Engineering Society Convention: 127,* Paper Number: 7879.

Woszczyk, W. (2003) Physical and Perceptual Considerations for High-resolution Audio. *Audio Engineering Society Convention: 115*, Paper Number: 5931

Yoshikawa, S., Noge, S., Ohsu, M., Toyama, S., Yanagawa, H., & Yamamoto, T. (1995). Sound-quality evaluation of 96-kHz sampling digital audio. *Audio Engineering Society Convention: 99*, Paper Number: 4112.

Zentner, M., Grandjean, D., & Scherer, K.R. (2008). Emotions evoked by the sound of music: Characterization, classification and measurement. *Emotion*, 8 (4), 494 – 521.