A Normative Survey on Auditory Perception and Semantic Descriptors of Musical Timbres

[1*]Anis Haron, [2] Wong Chee Onn and [3] Hew Soon Hin
[1,2,3] Faculty of Creative Multimedia, Multimedia University, 63100 Cyberjaya, Selangor,
email: [1*]1171402185 @student.mmu.edu.my, [2] cowong@mmu.edu.my, [3] shhew@mmu.edu.my

Date received: 29 April 2021
Date accepted: 24 September 2021
Date published: 30 November 2021

Abstract - *Timbre are commonly described using semantic descriptors such as 'dark', 'bright' and 'warm'. The use of such descriptors are useful and largely practiced by trained individuals in music related industries. Such descriptors are subjective as it could be interpreted differently by different individuals determined by factors such as training and exposure. Semantic descriptors also lacks granularity, in a sense that the descriptor does not indicate the amount or intensity of the description. A numerical representation for timbral description addressees these issues. Computational approach for numerical measure of timbre are at present under study by music technology researchers. Such studies requires a benchmarking process in order for viability tests. To provide a set of data that can be used for benchmarking, a survey on auditory perception and semantic descriptors of musical timbres were conducted. The conducted survey looks to find out if a general consensus can be observed for semantic description of musical timbres using a normative survey methodology. This article reviews the conducted survey, presenting the survey's approach, results and findings.*

Keywords: Data science, Tone colour, Perception, Semantic descriptors, Sound and music computing

1 Introduction

Studies of perception in relation to sound and music are historically focused on consonance and dissonance of intervals. Guthrie and Morrill have shown that the concept of consonance and pleasantness lead to similar responses. Guernsey argued that the perception of consonance is also conditioned by natural sensory processes, training, environment and musical context. This means musical familiarity contributes to the perception of what is considered to be consonant. For example, a subject with western music training might have a different perspective on consonance compared to a subject with classical Indian music training. Plomp and Levelt stated that for trained individuals such as musicians, ranking of interval for consonance differs from ordering in term of pleasantness. However, for the average individual or musically untrained subjects these are similar concepts as described by Guthrie and Morrill. These studies have greatly contributed towards understanding perception of consonance, dissonance and pleasantness of intervals and to a larger extend, contributes towards understanding timbres.

Descriptors used in audio are generally content-based descriptors, which include temporal, spectral, time-varying and harmonic descriptors (Caetano et al., 2019). These descriptors lend itself useful for research in music information retrieval with applications such as identification of musical instruments (Rodrigo et al., 2021). Semantic descriptor as an audio feature are perception-based descriptors, which requires human identification. Model for timbral qualities are developed based on semantic descriptors such as rumbling/low, soft/singing, brassy/metallic among others (Reymore & Huron, 2020). Such model finds application in characterizing musical instrument timbres and orchestration (Reymore, 2021). However, such use will be restricted to western instruments as the model relies on human identification of descriptors that was limited to western musical instruments. A recent study suggested that timbre brightness perception is primarily perceived by acoustical cues (Saitis & Siedenburg, 2020). Prompting us to observe the potential of using semantic descriptors for identification of timbral qualities commonly used in music production. Audio engineers use descriptors such as dark and bright among other descriptors to describe timbral qualities of sounds in music production processes such as equalisation (Stables et al., 2014).

This article reports on the findings of a conducted normative survey that looks to find out if a general consensus can be observed for semantic description of musical timbres. Responses that forms the majority and are within one standard deviation from the mean will be considered the normal. Identifying if a general consensus can be observed contributes to the research and development of a numerical measure of timbral descriptors, addressing the issue of granularity of description.

The article starts by covering the methodology of the conducted survey, followed by source of samples used, scope of the conducted survey and design of the survey's questionnaire. The following section of this article reports the results of the conducted survey, which is followed by findings and discussion sections.

2 Methodology

This study employs a normative survey research methodology seeking to answer the following question:

- Is there a general consensus for the identification and classification of a semantic descriptor for a given timbre?

Identification is defined as the ability to identify a semantic descriptor, while classification is defined as the ability to rank from most to least within the same semantic descriptor. Participants are required to listen to audio samples and provide the best descriptor for the particular audio sample. The survey was made available online for two months starting end of December 2020 and was disseminated via Multimedia University's student and faculty mailing lists, audio production online forums, and social media.

2.1 Source and scope

Audio samples used in this research was taken from freesound, a collaborative database of audio samples released under the Creative Commons licenses which allows for reuse. Audio samples in freesound database are paired with audio descriptor keyword tag that is provided by the database user whom have shared the audio sample to the database. Semantic audio feature extraction (SAFE) project by Semantic Audio Labs developed a suite of digital audio workstation plug-ins that allow users to both save and load semantic terms. Parameter settings for the plug-ins are available in a database. We focused our survey to four most commonly used user defined audio descriptors, obtained from the SAFE database. These descriptors are bright, dark, warm and thin.

Participants of the survey are reminded to judge only based on the timbre, tone colour or the perceived tone quality of the audio samples without taking into consideration its melody, rhythm, pitch, loudness and space. For each of the four identified commonly used audio descriptors, nine samples were selected based on the relevance ranking and downloaded from freesound database. Samples used in the survey were randomly selected from the ten downloaded samples for each of the four commonly used audiodescriptors. File names of audio samples used in the survey is hidden from survey respondents to avoid the possibility of association by file name.

2.2 Design and procedure

The survey starts with a basic demographic information section followed by five sections of questions. In the demographic section, participants will disclose their countryof residence, age group, musical training, industry affiliation and lastly, equipment used and environmental considerations while taking the survey. In section 1, participants were asked to choose the best audio descriptor from a list to answer which descriptor is deemed best fit for a group of audio samples. There are four groups in this section, with three audio samples in each group. Participants are allowed to provide the same descriptor for more than one group.

Sections 2, 3, 4 and 5, are categorised based on audio descriptors as provided in freesound. Audio samples in section 2 are samples described as 'warm', audio samples in section 3 described as 'dark', section 4 samples described as 'bright' and lastly section samples described as 'thin'. In each section from section 2 to section 5, there are four questions in total. The first three questions in sections 2 to 5 are multiple choice questions where participants were asked to listen to three pairs of audio sample and rank either option A or option B in each pair of samples as being brighter, darker, warmer or thinner than the other option. The last question in each section are ordinal questions, where participants were asked to listen to four audio samples and rank each audio sample from the most warm, bright, dark or thin to the least warm, bright dark or thin.

3 Results

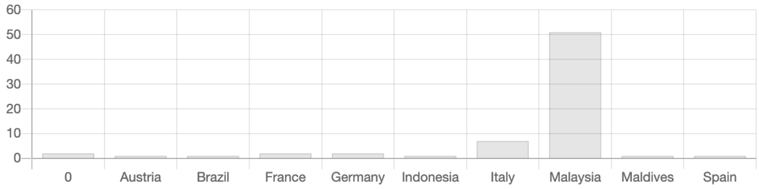

Figure 1: Demographic – Country

3.1 Demographics

As illustrated in figure 1, there are a total of 69 survey respondents originating from 9 different countries. The largest group of respondents are from Malaysia with 51 respondents. Out of the 69 respondents, 37 are below the age of 30 years old while the remaining 32 respondents are above the age of 30 years old, as shown in figure 2. As illustrated in figure 3, 34 respondents are not musically trained, while 35 respondents are musically trained. Out of the 35 musically trained respondents, 32 are engaged in music as a pastime activity or as a hobby while the remaining 3 respondents are professional musicians. As shown in figure 4, 59 of the respondents are not affiliated in any music related industry, while 10 respondents are affiliated in a music related industry. Figure 5 illustrates that out the 69 respondents, 23 respondents used headphones or earphones while taking the survey, 34 respondents did not use headphones or earphones but was in a quiet environment while taking the survey and the remaining 12 respondents are both not in a quiet space and not using headphones or earphones.

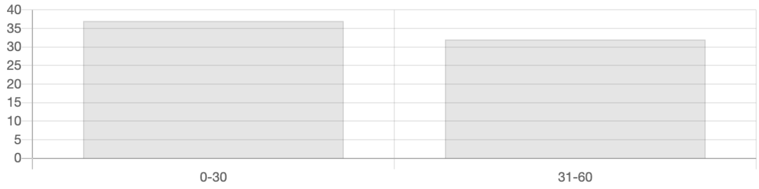

Figure 2: Demographic – Age group

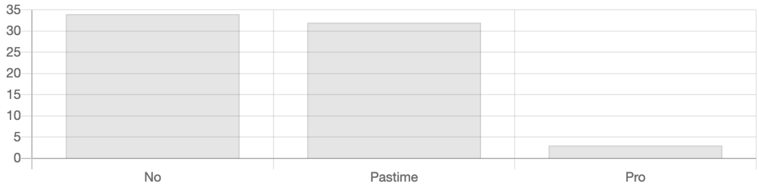

Figure 3: Demographic – Music training

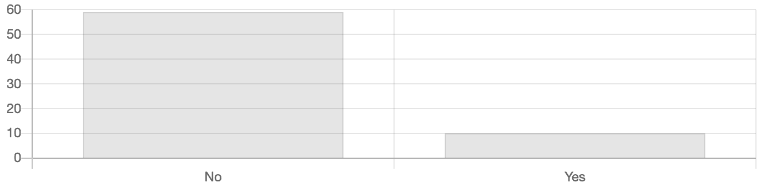

Figure 4: Demographic – Industry

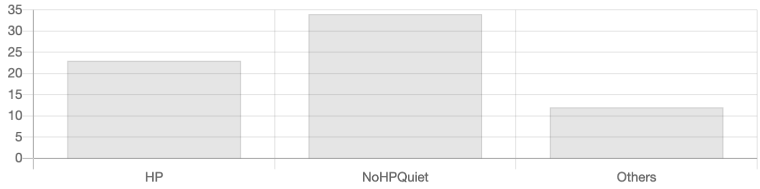

Figure 5: Demographic – Setup

Section 1: Identification of audio descriptors

In this section, four groups of three audio samples were provided. Survey participants were asked to choose the audio descriptor that is deemed best fit the group of audio samples.

Group A - Bright

Audio samples in this group are tagged ’bright’ in freesound database. Out of the 69 respondents, 45 respondents selected ’bright’ as being the best fit audio descriptor in this group. 13 respondents selected ’thin’, and the remaining 11 respondents selected ’warm’. No respondents selected ’dark’ for this group of audio samples. Figure 6 shows a plot of respondent responses.

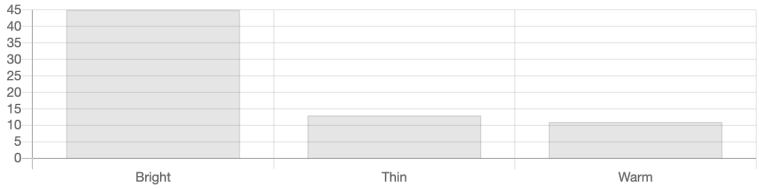

Figure 6: Section 1 - Bright

Group B - Dark

Audio samples in this group are tagged ’dark’ in freesound database. Out of the 69 respondents, 56 respondents selected ’dark’ as being the best fit audio descriptor in this group. 8 respondents selected ’warm’, 3 respondents selected ’thin’ and the remaining 2 respondents selected ’bright’ for this group of audio samples. Figure 7 shows a plot of respondent responses.

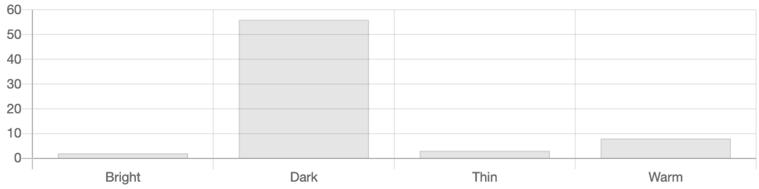

Figure 7: Section 1 – Dark

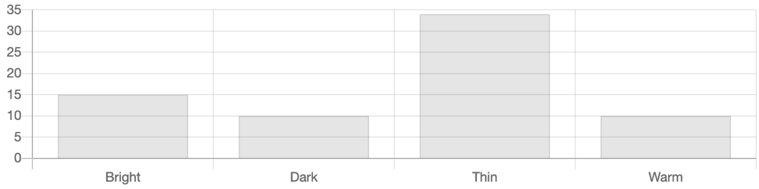

Figure 8: Section 1 - Thin

Group C - Thin

Audio samples in this group are tagged ’thin’ in freesound database. Out of the 69 respondents, 34 respondents selected ’thin’ as being the best fit audio descriptor in this group. 15 respondents selected ’bright’, while 10 respondents selected ’dark’ and ’warm’ respectively for this group of audio samples. Figure 8 shows a plot of respondent responses.

Group D - Warm

Audio samples in this group are tagged ’warm’ in freesound database. Out ofthe 69 respondents, 29 respondents selected ’warm’ as being the best fit audio descriptor in this group. 27 respondents selected ’dark’, 8 respondents selected ’bright’ and the remaining 5 respondents selected ’thin’ for this group of audio samples. Figure 9 shows a plot of respondent responses.

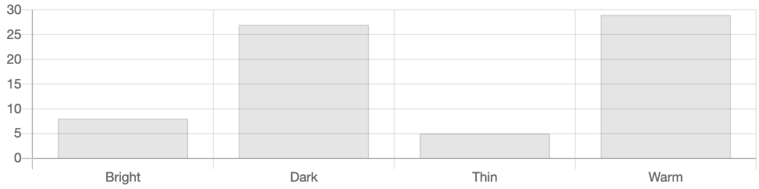

Figure 9: Section 1 – Warm

Section 2: Classification of audio samples with the same descriptor – Warm

In question 1, 2 and 3, survey participants were asked to choose which audio sample sounds warmer between option A and option B. In question 1, 47 respondents ranked option A (warm09.wav) as warmer than option B (warm03.wav), as shown in figure 10. In question 2, 46 respondents ranked option A (warm06.wav) as warmer than option B (warm02.wav), as shown in figure 11. In question 3, 43 respondents ranked option A (warm07.wav) as warmer than option B (warm05.wav), as shown in figure 12. In question 4, survey participants were asked to listen to four audio samples and rank them from most warm to least warm. Majority of participants ranked clip A (warm04.wav) as most warm, followed by clip D (warm07.wav) as warm, clip B (warm01.wav) as slightly warm, and lastly clip C (warm08.wav) as least warm, as shown in figure 13.

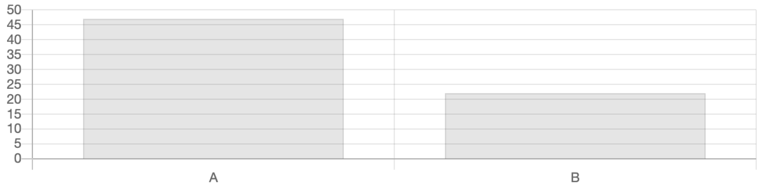

Figure 10: Section 2 – Question 1

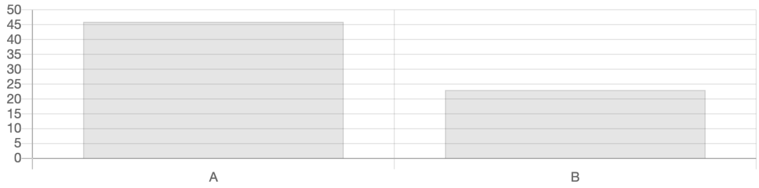

Figure 11: Section 2 – Question 2

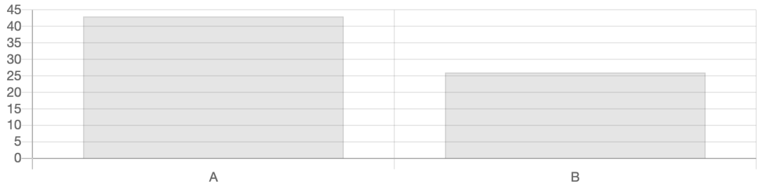

Figure 12: Section 2 – Question 3

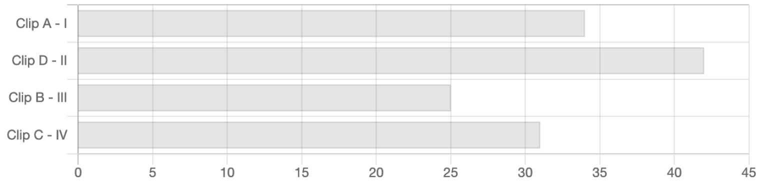

Figure 13: Section 2 – Question 4

Section 3: Classification of audio samples with the same descriptor – Dark

In question 1, 2 and 3, survey participants were asked to choose which audio sample sounds darker between option A and option B. In question 1, 43 respondents ranked option B (dark01.wav) as darker than option A (dark03.wav), as shown in figure 14. In question 2, 45 respondents ranked option B (dark04.wav) as darker than option A (dark06.wav), as shown in figure 15. In question 3, 35 respondents ranked option B (dark08.wav) as darker than option A (dark05.wav), as shown in figure 16. In question 4, survey participants were asked to listen to four audio samples and rank them from most dark to least dark. Majority of participants ranked clip C (dark07.wav) as most dark, followed by clip A (dark02.wav) as dark, clip D (dark06.wav) as slightly dark, and lastly clip B (dark09.wav) as least dark, as shown in figure 17.

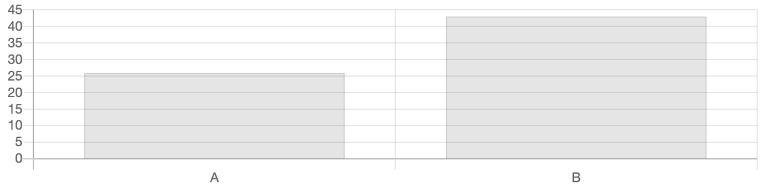

Figure 14: Section 3 – Question 1

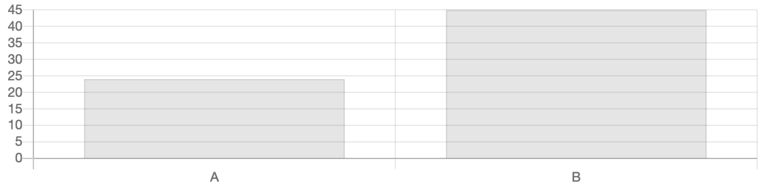

Figure 15: Section 3 – Question 2

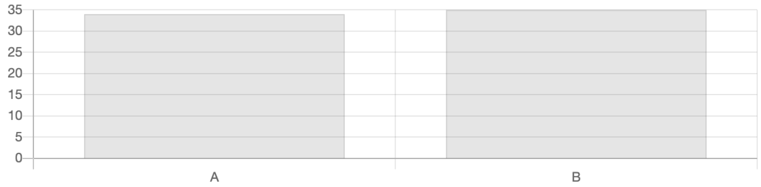

Figure 16: Section 3 – Question 3

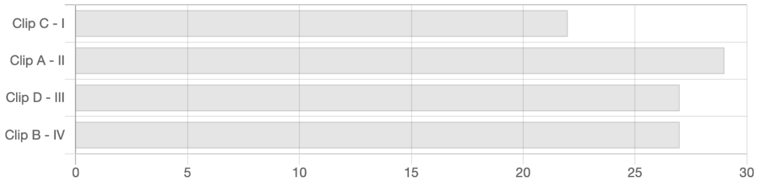

Figure 17: Section 3 – Question 4

Section 4: Classification of audio samples with the same descriptor – Bright

In question 1, 2 and 3, survey participants were asked to choose which audio sample sounds brighter between option A and option B. In question 1, 54 respondents ranked option A (bright03.wav) as brighter than option B (bright06.wav), as shown in figure 18. In question 2, 55 respondents ranked option B (bright01.wav) as brighter than option A (bright08.wav), as shown in figure 19. In question 3, 45 respondents ranked option B (bright09.wav) as brighter than option A (bright04.wav), as shown in figure 20. In question 4, survey participants were asked to listen to four audio samples and rank them from most bright to least bright. Majority of participants ranked clip D (bright08.wav) as most bright, followed by clip C (bright07.wav) as bright, clip B (bright02.wav) as slightly bright, and lastly clip A (bright05.wav) as least bright, asshown in figure 21.

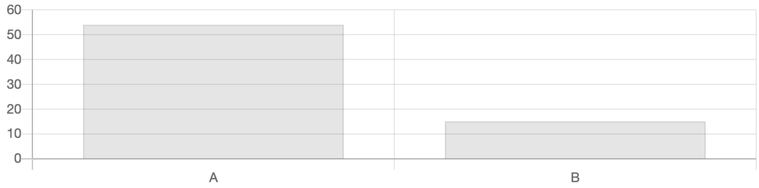

Figure 18: Section 4 – Question 1

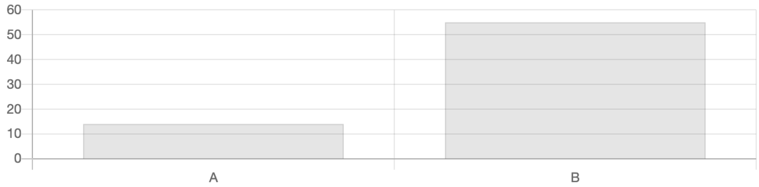

Figure 19: Section 4 – Question 2

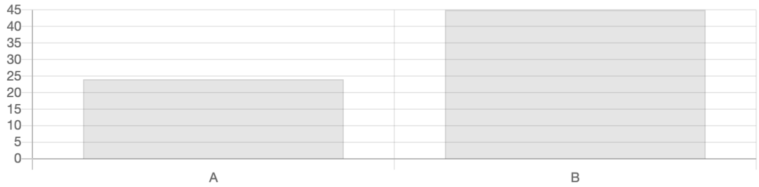

Figure 20: Section 4 – Question 3

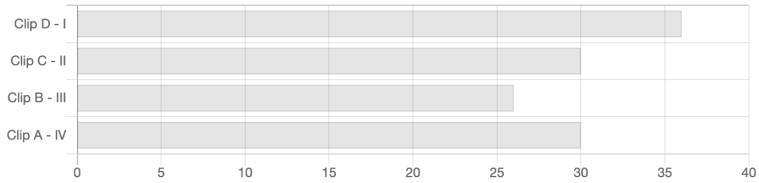

Figure 21: Section 4 – Question 4

Section 5: Classification of audio samples with the same descriptor – Thin

In question 1, 2 and 3, survey participants were asked to choose which audio sample sounds thinner between option A and option B. In question 1, 53 respondents ranked option B (thin02.wav) as thinner than option A (thin07.wav), as shown in figure 22. In question 2, 43 respondents ranked option A (thin04.wav) as thinner than option B (thin01.wav), as shown in figure 23. In question 3, 45 respondents ranked option B (thin09.wav) as thinner than option A (thin08.wav), as shown in figure 24. In question 4, survey participants were asked to listen to four audio samples and rank them from most thin to least thin. Majority of participants ranked clip B (thin03.wav) as most thin, followed by clip A (thin06.wav) as thin, clip C (thin05.wav) as slightly thin, and lastly clip D (thin01.wav) as least thin, as shown in figure25.

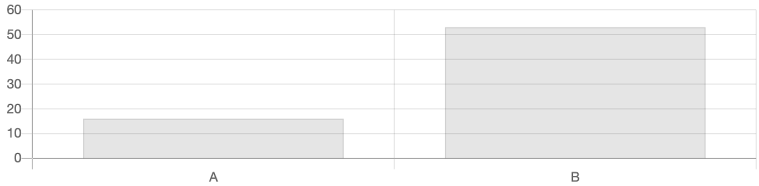

Figure 22: Section 5 – Question 1

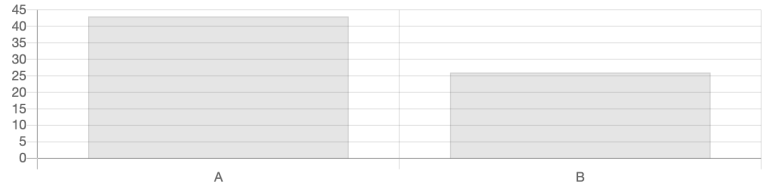

Figure 23: Section 5 – Question 2

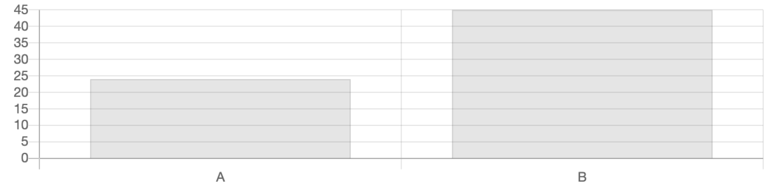

Figure 24: Section 5 – Question 3

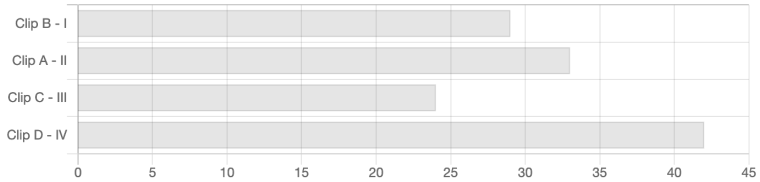

Figure 25: Section 5 – Question 4

3.2 Findings

To get an accuracy score for each respondent, we tabulated every response from each respondent and compared each response against the highest scored answer for the corresponding question. This means, if a respondent's answers are the same as the highest scored answer for every question, this respondent's accuracy score will be 100%. Respondents are then categorised into three accuracy categories based on this score - low (0-32%), average (33-65%) and high (66-100%). 16 respondents, which is 23% of the respondents were identified as high accuracy respondents, 51 respondents or 74% of respondents scored averagely, while 2 respondents or 3% of respondents were identified as low accuracy respondents.

Looking at the age group for high accuracy respondents, 44% of respondents are aged below 30 while 56% of respondents are aged above 30. As for the average accuracy respondents, 57% of respondents are aged 30 or below while 43% of them are aged above 30. While for the low accuracy respondents, 50% of respondent aged below 30 and the other 50% above the age of 30.

As for musical training of the respondents, 100% of respondents whom are classified as low accuracy respondents did not have any musical training. For average accuracy respondents, 45% of respondents did not have any musical training, 51% of respondents engage in music as a pastime activity, while the remaining 4% of respondents are professional musicians. As for the high accuracy respondents, 56% of respondents did not have any musical training, 38% engage in music as a pastime activity, and 6% of respondents are professional musicians. In total, the majority of the respondents did not have any musical training. However, a majority of average accuracy respondents engage in music as a pastime activity. These findings suggest that musical training is not an absolute criterion for higher accuracy responses, but only respondents without musical training scores with lower accuracy. Similar can be said for industry affiliation. Only a small percentageof respondents are affiliated to a music related industry, which suggest being industry affiliated is not a criterion for high accuracy responses, however, all low accuracy respondents are not affiliated to a music related industry.

For equipment and environmental consideration, 42% of all respondents used headphones or earphones, 48% of respondents did not but participated in the survey in a quiet environment, while the remaining 9% of respondents did not use headphones or earphones and are not in a quiet environment. We found no clear connection between equipment and environmental consideration with respondent's accuracy, suggesting the accuracy of responses are not dependent on equipment and environmental consideration.

In section 1, participants were asked to select an audio descriptor that are deemed to be the best fit for four groups of three audio samples. In group A, the three audio samples presented are samples tagged as ’bright’ in freesound. 65% of respondents selected ’bright’ as the best fit audio descriptor for group A. This shows that the majority of participants identified the samples correctly in group A. Similar results are observed for the remaining three groups. Group B consist of samples tagged as ’dark’, with 81% of respondents identifying it as ’dark’. Group C consists of samples tagged as ’thin’, with 49% of respondents identifying it as ’thin’. Lastly, group D consist of samples tagged as ’warm’, with 42% of respondents identifying it as ’warm’. From this, we found that consistently, a majority of respondents identified all four groups of samples correctly, as in, the majority of respondents consistently agree to the audio descriptor as it was tagged in freesound.

In section 2, for the first three questions, respondents ranked audio samples tagged as ’warm’ to distinguish which sample is warmer between three pairs of samples. In the final question, respondent rank four audio clips from most warm to least warm. The same procedure applies for section 3, 4 and 5 for different audio descriptor. For the first three questions in section 2, 68% of respondents ranked sample named ’warm09’ to be warmer than sample ’warm03’. 67% of respondents ranked sample ’warm06’ to be warmer than sample ’warm02’. 62% of respondents ranked sample ’warm07’ to bewarmer than sample ’warm05’. For the final question, a majority of participants ranked samples ’warm04’ (49%), ’warm07’ (61%), ’warm01’ (36%) and sample ’warm08’ (45%) from most warm to least warm respectively.

In section 3, for the first three questions, 62% of respondents ranked sample named ’dark01’ to be darker than sample ’dark03’. 65% of respondents ranked sample ’dark04’ to be darker than sample ’dark06’. 51% of respondents ranked sample ’dark08’ to be darker than sample ’dark05’. For the final question, a majority of participants ranked samples ’dark07’ (32%), ’dark02’ (42%), ’dark06’ (39%) and sample ’dark09’ (39%) from most dark to least dark respectively.

In section 4, for the first three questions, 78% of respondents ranked sample named ’bright03’ to be brighter than sample ’bright06’. 80% of respondents ranked sample ’bright08’ to be brighter than sample ’bright01’. 65% of respondents ranked sample ’bright09’ to be brighter than sample ’bright04’. For the final question, a majority of participants ranked samples ’bright08’ (52%), ’bright07’ (43%), ’bright02’ (38%) and sample ’bright05’ (43%) from most bright to least bright respectively.

Lastly in section 5, for the first three questions, 77% of respondents ranked sample named ’thin02’ to be thinner than sample ’thin07’. 62% of respondents ranked sample ’thin04’ to be thinner than sample ’thin01’. 65% of respondents ranked sample ’thin09’ to be thinner than sample ’thin08’. For the final question, a majority of participants ranked samples ’thin03’ (42%), ’thin06’ (48%), ’thin05’ (35%) and sample ’thin01’ (61%) from most thin to least thin respectively.

3.3 Discussion

Findings from this study have shown that there is a general consensus for the identification of an appropriate semantic audio descriptor for a given audio sample. In section 1 of the survey, the results are consistent though out all four semantic audio descriptors used in this study. Although with varying numbers, a majority of respondents correctly identified the appropriate semantic audio descriptor in all cases. Results in section 1 recorded a mean percentage of 59.25%, with standard deviation of 17.4%. Responses for descriptors ‘bright’, ‘thin’ and ‘warm’ are within one standard deviation from the mean. While for the descriptor ‘dark’, response was above one standard deviation from the mean, as shown in figure 26.

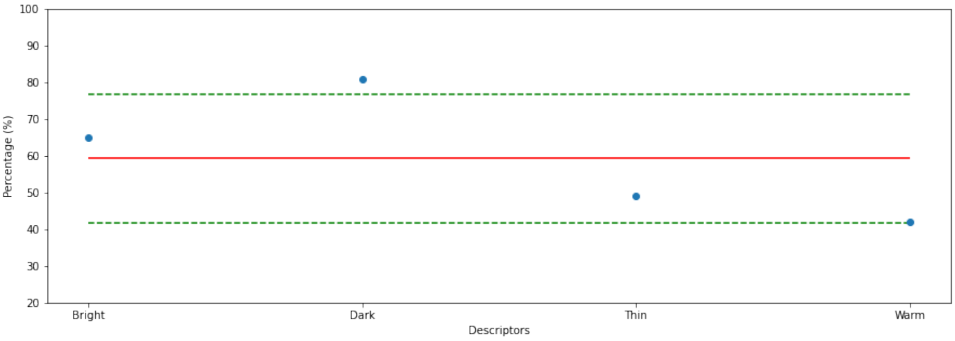

Figure 26: Survey results – Section 1

Similarly can be observed for classification of an appropriate semantic audio descriptors in the multiple choice questions (questions 1 to 3 in sections 2 to 5). A general consensus is present in almost all cases across the board with the largest majority recorded at 80%. Results in the multiple choice questions recorded a mean percentage of 66.83%, with standard deviation of 8.16%. Responses are within one standard deviation from the mean with the exception of the following, as shown in figure 27:

Section 3, question 3 records a response below one standard deviation.
Section 4, question 1 records a response above one standard deviation.
Section 4, question 2 records a response above one standard deviation.
Section 5, question 1 records a response above one standard deviation.

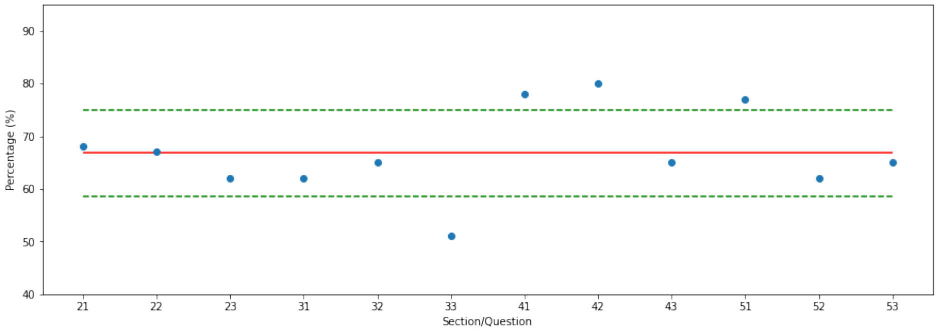

Figure 27: Survey results – Section 2, 3, 4 & 5 – Multiple choice questions

In the ordinal questions (question 4 in sections 2 to 5), the largest majority is recorded at 61%, while the smallest majority is recorded at 32%. Responses are grouped according to the ordinal ranks from highest ('most') to lowest ('least'). In the highest ordinal rank group ('most'), mean percentage are recorded at 43.75%, with a standard deviation of 8.88%. Response for the descriptor 'dark' recorded below one standard deviation from the mean with the remaining descriptors recording responses within one standard deviation as illustrated in figure 28.

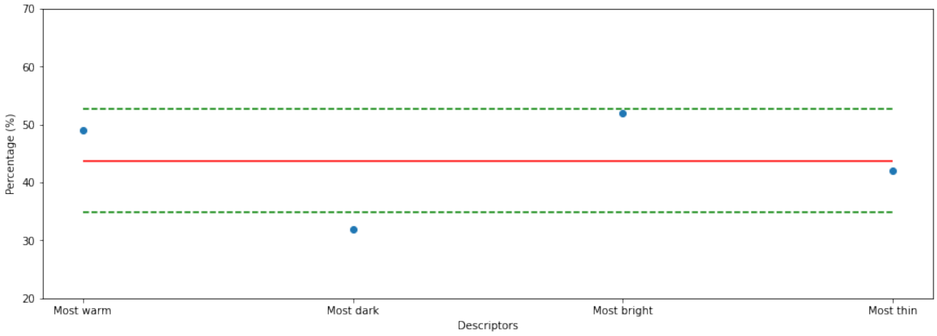

Figure 28: Survey results – Section 2, 3, 4 & 5 – Ordinal questions ('Most')

As for the second highest ordinal rank group, mean percentage are recorded at 48.5%, with a standard deviation of 8.74%. Response for the descriptor 'warm' recorded above one standard deviation from the mean with the remaining descriptors recording responses within one standard deviation as illustrated in figure 29.

While for the third highest ordinal rank group ('slightly'), mean percentage are recorded at 37%, with a relatively smaller standard deviation of 1.83%. Response for the descriptor 'dark' recorded just above one standard deviation from the mean, while for the descriptor 'thin' recorded just below one standard deviation from the mean and the remaining descriptors recording responses within one standard deviation as illustrated in figure 30.

Finally for the lowest ordinal rank group ('least'), mean percentage are recorded at 47%, with a standard deviation of 9.67%. Response for the descriptor 'thin' recorded above one standard deviation from the mean with the remaining descriptors recording responses within one standard deviation as illustrated in figure 31.

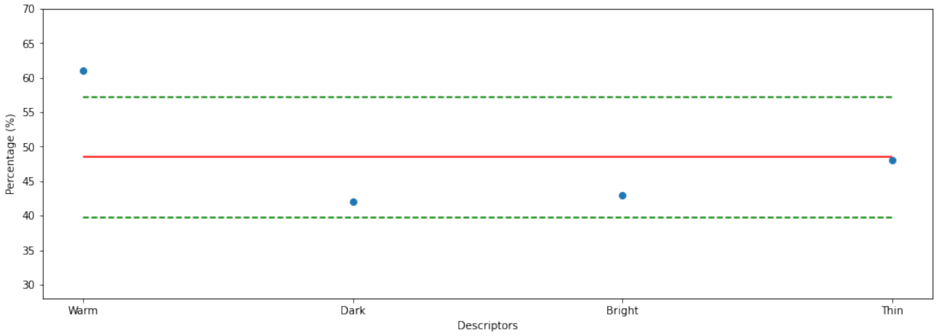

Figure 29: Survey results – Section 2, 3, 4 & 5 – Ordinal questions ('Neutral')

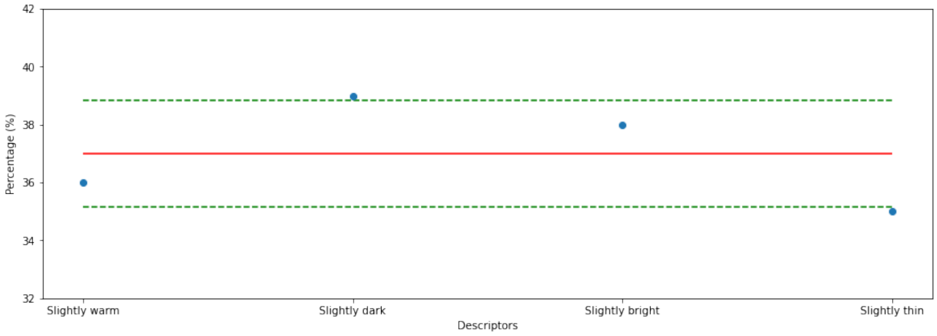

Figure 30: Survey results – Section 2, 3, 4 & 5 – Ordinal questions ('Slightly')

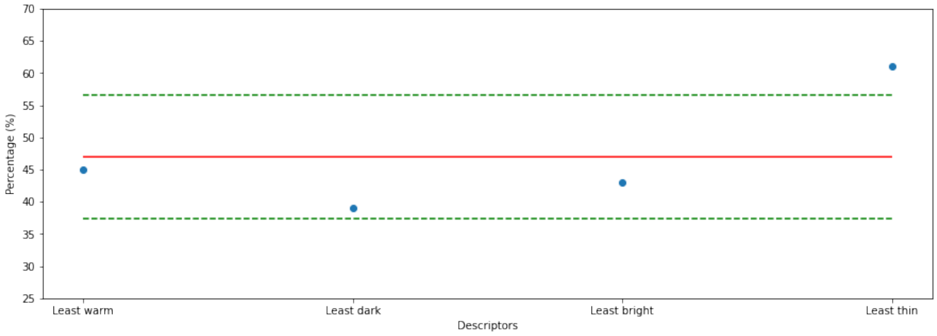

Figure 31: Survey results – Section 2, 3, 4 & 5 – Ordinal questions ('Least')

This study have also shown that musical familiarity is not a major determining factor which could affect the ability to identify and classify semantic audio descriptors for a given audio sample. Respondents of the survey largely scored with average accuracy (74%), with respondents in this category having close to an even spread between respondents with musical training (55%) and without musical training (45%). As for music industry related affiliation, 88% of respondents in this category are not industry affiliated. Similarly for survey respondents with high accuracy scores (23%), respondents in this category too have close to an even spread between respondents with musical training (44%) and without musical training (56%). As for music industry related affiliation, 75% of respondents in this category are not industry affiliated. Very few respondents scored with low accuracy(3%). However, since all low accuracy scorers have no musical training and no music industry related affiliation, it can be inferred that while musical familiarity is not a major determining factor, chances to score with below average accuracy could be higher for respondents without musical familiarity.

4 Conclusions

Survey results from section 1 of the survey which focuses on the identification have shown that given a set of audio samples, respondents are able to identify the appropriate semantic descriptor for all cases. Results from section 1 also have shown that a larger percentage of respondents are able to correctly identify the semantic descriptor for 'bright' and 'dark' descriptors. As for 'thin' and 'warm' descriptors, respondents correctly identified the descriptors with a lower than the mean percentage, albeit still within one standard deviation from the mean. As for the classification of semantic descriptors addressed in sections 2, 3, 4 and 5 of the survey, the process is split into classification of two audio samples addressed with three multiple choice questions (questions 1, 2 and 3) and one ordinal question (question 4) to classify four audio samples. Survey results shown that for multiple choice questions responses are largely within one standard deviation from the mean with three instances being above one standard deviation from the mean and one instance being below one standard deviation from the mean. From this result we can infer that for instances that are above one standard deviation from the mean are classifications of higher assurance, while for the instance below one standard deviation from the mean are classification of weaker assurance. As for survey results on ordinal questions, the mean percentage for each ordinal rank are lower compared to the mean percentage for multiple choice questions, which infers classifications of weaker assurance as a whole. However, in most instances, responses are largely within one standard deviation from the mean with some instances being just above and just below one standard deviation. This results suggests that respondents might find it easier to classify descriptors when presented with only two audio samples to be compared against and finds it more difficult to classify when a larger number of options are presented. To conclude, results from the conducted survey have demonstrated that in most cases, a clear inclination towards a general consensus can be observed in both identification and classification of a semantic descriptor for a given timbre.

Acknowledgements

The authors would like to express appreciation for the support of the sponsor, MMU Graduate Research Assistant Scheme (MMUI/180023.02) without which this research would not have been possible.

References

Caetano, M., Saitis, C., & Siedenburg, K. (2019). Audio Content Descriptors of Timbre. Timbre: Acoustics, Perception, and Cognition. Springer Handbook of Auditory Research, 69, 297–333. https://doi.org/10.1007/978-3-030-14832-4_11

Rodrigo, W. U. D., Ratnayake, H. U. W., & Premaratne, I. A. (2021). Identification of Music Instruments from a Music Audio File. *Lecture Notes in Networks and Systems*, 335–352. https://doi.org/10.1007/978-981-33-4355-9_26

Reymore, L., & Huron, D. (2020). Using auditory imagery tasks to map the cognitive linguistic dimensions of musical instrument timbre qualia. *Psychomusicology: Music, Mind, and Brain*, *30*(3), 124–144. https://doi.org/10.1037/pmu0000263

Reymore, L. (2021). Characterizing prototypical musical instrument timbres with Timbre Trait Profiles. *Musicae Scientiae*, 102986492110015. https://doi.org/10.1177/10298649211001523

Stables, R., Enderby, S., De Man, B., Fazekas, G., & Reiss, J. D. (2014). Safe: A System for Extraction and Retrieval of Semantic Audio Descriptors. In The International Society of Music Information Retrieval. ISMIR2014

Saitis, C., & Siedenburg, K. (2020). Brightness perception for musical instrument sounds: Relation to timbre dissimilarity and source-cause categories. *The Journal of the Acoustical Society of America*, *148*(4), 2256–2266. https://doi.org/10.1121/10.0002275