Bridging disparity: An attempt to train consumers in music technology using 'audio descriptors'

Harigopal Sasikumar

IMSE
King Mogkut's Institute of Technology Latkrabang
Bangkok, Thailand
harigopal.sa@kmitl.ac.th

***Abstract*—High end audio systems with premium price tags have been invading home spaces for a while and heavily data compressed low quality files are played back through the system. Educating consumers on audio technology is proving to be relevant for better product judgement and audio quality appreciation. This particular study is conducted to obtain the response of an untrained listener when exposed to the technical aspects of audio and take the first step towards training the consumer in audio technology using 'audio descriptors' found in the EQ(equalizer) plugin of commonly used DAW (Digital Audio Workstation) from Apple called Logic Pro. Since frequency is the most relevant aspect of sound, the intention was to train the technical aspect of these terms to the participants with the help of audio spectrum and human hearing range. Statistical analysis of the listening tests shows the possibilities of bridging the gap between the audio technology and consumers is very likely even though grasping advanced and complex technologies requires ear training and practise.**

Keywords—Descriptors, Consumers, EQ

I. INTRODUCTION

Audio engineers, programmers and manufacturers are considered as experts who can manipulate what they hear as well as capable of defining it in scientific terms. Does this factor of expertise in audio terminologies generate a gap between the audio technology and consumers?. Convenience created by the data compression is exploited by online audio/video streaming platforms and becoming a threat to audio quality and music appreciation. This indicates the necessity in educating consumers in audio technology, terminologies and various production methods employed[13]. While describing the quality of audio equipment such as loud speakers etc. people tend to use various adjectives such as "soft," "dark," "bright" and so on. This can create confusion regarding the judgment of quality. Considering the range and broadness of aspects covered it could be useful for the consumers if a set of audio descriptors which is more relatable to technical aspects is developed[2].

When looking from a different perspective, consumers can be trained to have a better understanding about audio processing which creates a wider market as well as space for advanced technologies and devices within current user trends. Study by Francis Rumsey points out the need for clarity in defining subjective aspects of audio. Rumsey also suggests three categories to be considered when looking at audio quality; technical, spatial and timbral qualities[12]. Equipment such as amplifiers, Loudspeakers, headphones are described often in terms of Impedance, frequency response etc. for not so technically sophisticated users this information is of no use. Relation between perception of reproduced sound quality and the physical parameters is unclear so that even a professional sound engineer cannot make use of this technical information[2].

The aim of the study is to assess whether a consumer could be trained on the aspect of equalization with respect to a suitable descriptor. Equalization is only one basic aspect of a larger spectrum of parameters. The motivation for choosing EQ's for the study is due to the contribution of frequency in shaping a sound. More specific aspects such as compression or time based effects could be introduced as further steps if this study is effective. For this particular experiment audio stimuli were a score for a film written by an Indian composer which is not commercially released but was allowed to be used in the experiment. The length of the piece was equally divided for three sections of the project. The piece itself is divided into three parts structurally. This particular piece is not claimed to be an ideal piece and thus leads to the question of variation in perception according to the musicality or genre of the music which should be further investigated.

II. LITERATURE REVIEW

A. Consumer trends

Digital broadcast methods in modern days have a tendency to reduce audio quality for easy download and compatibility with all reproduction systems[11]. Arrival of MP3 with compressed data structure is supposedly the core of the reduction in quality. The modern technology allows audio to be reproduced approximately up to 25000 kbps but eventually ends up downloaded in low quality MP3 (128-320 kbps) digitally compressed from the internet. The amount of music listened might have an improvement but the quality of the audio is affected. The easy and fast access to MP3 on the internet ruins the order of the songs in the album. Consumers tend to stream low quality music using streaming websites like YouTube which reduces the audio quality further for faster streaming. One important factor here is about response from the audio industry regarding the acceptance of these trends or creating awareness about the reduced quality of MP3 and other compressed formats[11]. Necessity of educating listeners about the importance of quality and distribution methods is increasing. In a series of studies conducted by Olive on the performance and preferences of untrained listeners' vs trained listeners shows that the trained listeners were more consistent in their ratings as compared to the untrained listeners. This helps to understand the importance of listeners being trained on audio parameters. It

978-1-7281-9978-8/20/$31.00 ©2020 IEEE

can provide accuracy in audio research as well as improve quality appreciation[9][10].

Unlike Early home hi-fi systems with adjustable EQ up to seven frequency bands these days' home systems come with preset EQ curves based on the music genre. Some devices come with loudness options and bass boost. These loudness buttons were introduced based on the conclusion that consumers do not understand the concept of equalization. This shows the dumbing down of the listeners and the pre-existing assumption that the consumer doesn't understand the technicalities and terminologies to benefit from. The preset EQ such as 'Hard rock', 'classical' or 'Hip-hop' could be found in some systems. A better approach would be to present options for increasing 'warmth' or reducing 'brightness' and further more like reducing 'standing waves' for a particular room can be done using simple adjustment of EQ (also can be done using modulation in musical harmonic content) [14]. A more scientific study of these descriptors should be carried out for a clear understanding enabling to educate the consumer to benefit from such advanced technologies and systems[13]. To bridge the gap a set of audio terminologies and descriptors should be universally accepted as a common audio language. Commercial viability of the home hi-fi systems will improve with an improved appreciation in quality of audio.

B. Audio Descriptors

Terminologies to refer to timbral qualities often come from emotional context, images, metaphors and mostly are non-scientific. The study by Rumsey suggests the practise of this rare correlation could be beneficial and certainly possible[12]. The reliability of these terms can be improved contributing a hike in developing a universal language with specific audio descriptors. The specific terms selected to study the frequency contribution and training for the test are 'warm', 'bright', 'dark' and 'telephone'[13]. Timbral terminologies can be analysed and related to quantifiable parameters for scientific study as similar to Loudness, Pitch, and Reverberation. which are obvious quantities. A chart was defined for certain timbral terms which relates to the variations in their presence across the spectrum[7]. Timbral adjectives were used to address certain characteristics of a sound and properties using a single word, but as the trend changed the usage of the descriptor has become convenient for the user to avoid being specific at parameters or technical attributes referred to in the particular sound.

The use of thermionic valves in the amplifiers (mostly vintage) originally led to the description 'Warmth' since it glows and let's heat off. Which is often used to describe high quality audio.

− boost in the 200-500Hz of the sound spectrum

− Dip in the mid-range (2-6 kHz) of the audio spectrum (2.5 – 6 kHz)

− non-linear compression similar to analogue audio (tape)

−Slight harmonic similar to analogue(thermionic) [5][6][7].

Similarly, it should be possible to define technical characteristics relating to other timbral adjectives. 'Warm' and 'bright' were the two highest ranked terms in a dataset of 210 unique adjectives. Secondly, the two terms are deemed to be sufficiently different enough to form an audible timbral variation in low dimensional space. They have relatively dissimilar timbral profiles, with brightness widely accepted to be highly correlated with the signal's spectral centroid, and warmth often attributed to the ratio of the first 3 harmonics to the remaining harmonic partials in the magnitude LTAS (Long-term-average spectrum)[16]. The term 'dark' is having similar attributes to warm EQ but mostly associated with the musicality of the piece. But 'dark' is also highly ranked in the previously mentioned set of terms. These are the reasons for choosing 'warm', 'bright' and 'dark'. Telephone EQ was chosen because of wide usage of the technique as well as gives a better understanding due to drastic difference in frequency content.

In a study conducted by Bruno Fazenda and Alex Wilson, subjects were asked to listen to music samples and rate the overall quality of the audio along with suitable word descriptions for the samples describing quality[15]. Results show that the description varies with respect to expertise of the participant and other features. Expert listeners used a smaller set of terms which shows the contribution of the technical aspects of the particular sample. No expert listeners used words from more emotional context which resulted in a large set of terms which have less or no similarities. Timbral adjectives usually associate with sensations like touch('warmth'), taste('sweet') etc. In some cases, relating the sound to images. The term 'bright' is supposed to have higher energy in the high mid-range of frequencies or attenuation in low mid-range[7].

As per the survey conducted in the UK, US and Japan, a large number of citizens innovate and modify products for their needs. The survey also shows that the innovators were highly educated or skilled and mostly male. This shows the necessity of extending and exposing audio technology education to common citizens and female consumers[3].

III. TRAINING, TESTING AND ANALYSIS

For an accurate idea of whether the listeners could be trained or not, various parameters have to be studied such as compressed audio, time based effects etc. For this particular experiment only EQ is considered. The audio was played on an Apple MacBook Pro using Ableton Live 9 software Subjects were trained and tested individually. Training includes explaining human hearing range, what is frequency and all the EQ descriptors separately. The EQ curve was presented to the subjects and explained what happens with the curves. The test setup consists of Laptop, Focusrite Scarlett Audio Interface and Sennheiser HD 280 Pro Headphones. The level was set to -6 Db on the channel fader. Each subject had to go through one training session and one test session. 16 clips were auditioned in one test session. 20 participants took part in the test, 11 male subjects and 9 female subjects. The same musical piece was used for all subjects.

A. Presets used

Logic Pro X EQ section is used for training. After an intense research through various mixing plugins such as waves, sound toys etc. the most convincing EQ curves were found in the Logic pro X channel EQ plugin which comes within the software. Logic Pro is the one of the most used software in the audio industry. Interface of the EQ is simple and friendly. It is a fully parametric EQ and includes various presets with audio descriptors.

Make warmer as seen in Fig. 1 has a boost of 3dB towards the bass end and attenuation of 5dB towards the high end as seen in the image. It also has a 5dB boost between 100 to 200 Hz as well as 300 to 400 Hz. A dip at 1Khz is also done for this particular EQ. Fig. 2

Add brightness is one of the widely used audio descriptors in audio production. It is also arguably the most technical term because boost in the higher frequency range enhances instruments such as cymbals, higher octave violins etc. As the image shows there is a 5-7 dB boost above 2000 Hz.

Make darker Fig. 3 is almost similar to Warm as per the EQ settings but with no extra bumps. It will be difficult for the subject to identify between warm and bright while the other two are comparatively easier to identify. This can give an understanding about the technical grasping power of the subject.

Telephone EQ is widely used in music production for 'boxy' or 'phony' effects. It is normally referred to the frequency that is being heard from the telephone receiver. As it can be seen in Fig. 4 there are significant cuts below and above mid and high mid frequency range.

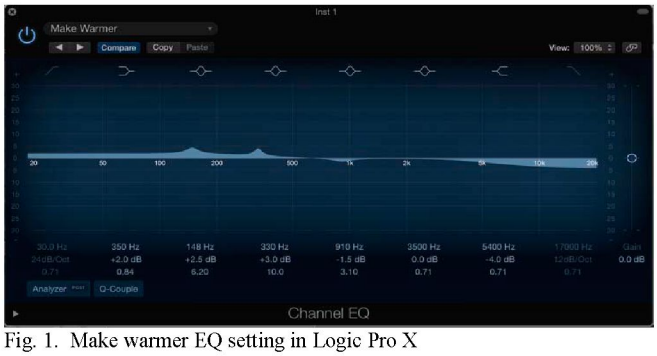

Fig. 1. Make warmer EQ setting in Logic Pro X

Fig. 2. Bright EQ setting in Logic Pro X

Fig..3. Dark EQ setting in Logic pro X

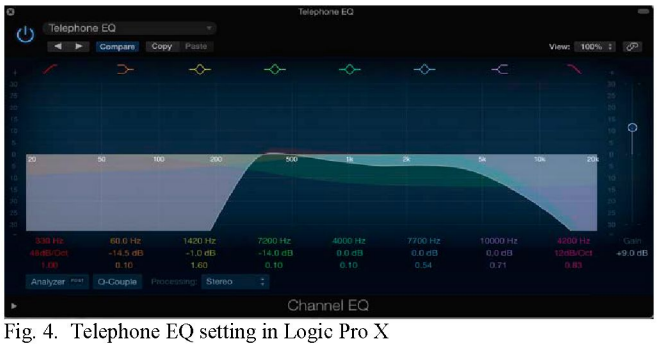

Fig. 4. Telephone EQ setting in Logic Pro X

B. *Listening test*

The subjects were provided with 16 clips of part 3 of the same piece of music. Four different Equalizations were applied. Four clips per EQ. All the clips are arranged randomly in an Ableton clip view window as shown in Fig. 5.

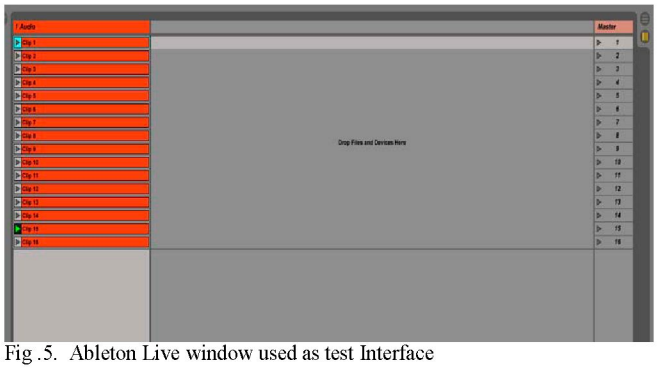
Fig .5. Ableton Live window used as test Interface

C. *Data Analysis*

Binomial testing is the employed analysis method for calculating the probability of guessing in multiple choice questions on every subject for each EQ setting. This value can determine the number of subjects successful in choosing the correct description without guessing. Conducting a hearing test before training could be Ideal to understand subjects' hearing. It was not done in this case.

One common use of the binomial test is in the case where the null hypothesis is that it has two equal possible outcomes (such as a toss of coin). The trials must be independent from one another otherwise it will become a hypergeometric distribution. Binomial test is commonly employed to find the success or guess rate (hypothesis) from n trails with k success. Where p is the probability.

16 clips with four different EQ settings 'warm', 'bright', 'dark' and 'telephone' were presented to the participant in a random order. Subjects were asked to pick one in four options. Four clips have the same EQ so the individual expected result for each EQ overall is 20 multiplied by 4.

Equation for binomial test:

$$P_n(x) = C(n, x)p^x q^{n-x} = n!/x!(n-x)! * p^x q^{n-x} \quad (1)$$

x = Number of success

p = 0.25 1/4th chance of guessing, q = 1 – p .

n= total number of trails

significance level for the test is 0.05 or 5%

D. Binomial test result for Individual EQ and Overall

Null hypothesis H_0 : Subjects are guessing.

TABLE I. BINOMIAL TEST RESULT

	n	x	p	P	H_0
Bright	80	45	0.25	P<0.000001	Rejected
Warm	80	35	0.25	P=0.000117	Rejected
Dark	80	30	0.25	P=0.00436	Rejected
Telephone	80	80	0.25	P is undefined	Rejected
Overall	320	232	0.025	P<0.000001	Rejected

All the individual EQ's and Overall response are run through the binomial test using equation(1) and the hypothesis H_0 is rejected in all cases which shows the subjects were not guessing. The subjects got over 50% correct answers for bright and telephone EQ's but less than 50% for warm and dark. It is obvious that recognizing tougher EQ curves are impossible with such short training. The overall testing shows it is possible for consumers to grasp music technology on a basic level.

IV. CONCLUSION

20 subjects were given training on 4 different types of EQ's and were asked to identify them in a test followed. The response from the subjects were positive and a general curiosity was visible during training and testing which is positive. The test results were analysed using binomial tests to see if subjects are guessing but result shows otherwise. As the result shows the possibilities of bridging the gap between the audio industry and consumers is very likely. The feedback back from the subjects reveals the developing interest in music technology. Indulging day to day users into innovation areas would help to improve audio quality appreciation. Extensive research and education across various platforms can help develop a set of audio descriptors as a universal language. The individual subject analysis shows that the subjects could easily identify Telephone EQ since it is having a heavy attenuation on the high and low end of the audio spectrum. 16 out of 20 participants correctly answered 'bright' EQ which shows the training is very likely to be successful. The other concerning factor is that the description used are from two extreme contexts but the spectral content is almost the same. There is no available literature based on what scientific ideology the EQ presets in Logic Pro is created. This can be very confusing for the content producer in the first place and calls for a study about music producers and their understanding of technology. This study is a step towards further enhanced researches. As mention earlier this study lacks hearing analysis of the subjects and developing a standalone platform for each technology (EQ in this case) can be beneficial for training and testing. As an immediate further step a study using actual consumer products would be beneficial for understanding the grasping capabilities of the participant and Helping the consumer understand and navigate through the frequent evolution of technology.

ACKNOWLEDGMENT

I am grateful to all of those who help me complete this study. I would like to thank each of the participants attending the training and being part of this experiment. Mostly I would like to thank Dr. Bruno Fazenda, for supervising and guiding me during the course of this study. As my teacher and mentor, he has taught me more than I could ever give him credit for here. He has shown me, by his example, what a good scientist (and person) should be.

REFERENCES

[1] S. Bech and N. Zacharov, "Perceptual audio Evaluation-Theory, method and application", John Wiley & Sons, 2007.

[2] A. Gabrielsson, and H. Sjögren, "Perceived sound quality of sound reproducing systems", The Journal of the Acoustical Society of America, 1979, pp. 1019-1033.

[3] E. P. Helzner, J. A. Cauley, S. R. Pratt, S. R. Wiśniewski, J. M. Zmuda, E. O. Talbott, ... and F.A. Tylavsky, "Race and sex differences in age related hearing loss: The Health, Aging and Body Composition Study", Journal of the American Geriatrics Society, 2005, pp. 2119-2127.

[4] Hippel, E. von, Ogawa, Susumu, Jong and P. J. Jeroen, "The Age of the Consumer-Innovator", MIT Sloan Management Review 53.1, 2011.

[5] Hood, J. L., "Valve and Transistor Amplifiers", Oxford, Newnes Publishing, 1997.

[6] Hood, J. L., "Audio Electronics", Oxford, Newnes Publishing. 1999.

[7] Katz, B., "Mastering Audio", Focal Press, 2002.

[8] Norušis, M. J., "*SPSS 14.0 guide to data analysis*", Upper Saddle River, NJ: Prentice Hall, 2006.

[9] Olive, Sean E., "Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: a case study", Journal of the Audio Engineering Society vol 51.9, pp. 806-825, 2003.

[10] Olive, Sean E., "Part 2- Differences in Performance of Trained Versus Untrained Listeners", Audio Musings, December 27, 2008.

[11] Owen, O., Tech Angst, Future Music, Future Publishing Ltd, pp. 13, 2006.

[12] Rumsey, F., "Psychoacoustics of Sound Quality", Proceedings of the Art of Record Production Conference, London, 2005.

[13] Toulson, E. R., "A need for universal definitions of audio terminologies and improved knowledge transfer to the audio consumer" In Presented at The Art of Record Production Conference, *Vol.* 8, pp. 10, 2006.

[14] White, P., "Creative Recording – part one, effects and processors", 2nd Edition, London, Sanctuary, 2002.

[15] Wilson, A., & Fazenda, B. M., "A lexicon of audio quality", In Proceedings of the 9th Triennial conference of the European Society for the Cognitive Sciences of Music (ESCOM 2015), Manchester, UK, August 2015.

[16] Stasis, S., Stables, R., & Hockman, J., "A model for adaptive reduced-dimensionality equalization", In Proceedings of the 18th International Conference on Digital Audio Effects, Trondheim, Norway, *Vol.* 30, December 2015.