IMPACT OF PERSONALIZED EQUALIZATION CURVES ON MUSIC QUALITY IN DICHOTIC LISTENING

Duo Zhang

MaxLinear Inc.

Carlsbad, CA, USA 92011

dzhangdustin@gmail.com

Hongmei Xia

Hubei Zhong Shan Hospital

26 Zhongshan Ave., Qiaokou District, Wuhan, Hubei, China 430033

xhmhongmei@gmail.com

Tiffany Chua

Henry Samueli School of Engineering
University of California, Irvine

Irvine, CA, USA 92697

tchua@uci.edu

Gerald A. Maguire

School of Medicine
University of California, Irvine

Orange, CA, USA 92868

gamaguir@uci.edu

David Franklin

School of Medicine
University of California, Irvine

Orange, CA, USA 92868

dlfrankl@uci.edu

Daniel Huang

Mophie Inc.
2850 Red Hill Avenue, Suite 128

Santa Ana, CA, USA 92705

daniel@mophie.com

Hung Tran

AuralWare LLC.

Rancho Santa Margarita, CA 92688

htntran@htmind.com

Hongbin Chen

Nurotron Inc.

Irvine, CA, USA 92018

hchen@nurotron.com

ABSTRACT

This paper investigated the impact of personalized equalization (EQ) on music quality. A pair of personalized EQ curves was found for each listener in a double-reference psychoacoustic test. Original high-fidelity music and music equalized by the pair of personalized EQ curves were randomly presented to listeners who were instructed to rate music quality. Statistical analysis showed that personally equalized music provided significantly higher music quality than original music.

1. INTRODUCTION

Equalization (EQ) is widely used in music players, studios, movie theatres, car stereo systems, home theatres, musical performances, etc. Equalization changes the amount of energy in different frequency bands, changing the timbre and character of music that a listener perceives, such as making drums sound more resonant, making a singer's voice more sensational, enhancing clarity, etc. Equalization is an integral part of users' interactive listening experience.

Traditionally, music equalization is performed in analog domain by time-domain filtering or frequency-domain manipulation using analog electronic components, resulting in inaccurate and inflexible equalization curves. The advent of digital technology has enabled high quality signal processing. Digital equalization filters can be designed directly in the digital domain by exact derivation of stable minimum-phase digital equalizer without resorting to analog prototyping [1]. This shortens design procedures, because traditionally digital design begins with analog design followed by application of the bilinear transform. Digital equalization techniques such as multi-band linear phase equalization are commonly used to provide linear phase and uniform frequency response [2].

Digital equalization has many benefits over traditional analog equalization. It allows the creation of equalization systems with greater flexibility and configurability, resulting in superior sound quality. Using stereo instead of mono equalization improves three-dimensional (3D) sound reproduction [3]. Using psychoacoustically motivated equalization filters instead of traditional root-mean-square (RMS) averaging equalization filters result in better performance [4]. Psychoacoustically motivated filters take into account the perceptual model of auditory system, and can be implemented with low computational complexity using fixed-pole parallel low-order filters [5]. The low computational-complexity filter designs make implementation feasible on a broad range of music software and hardware platforms.

The effect of digital equalization on music quality can be evaluated using modern psychoacoustic test methods that systematically evaluate individual psychometric functions. In general, adaptive methods are a better choice than constant-stimulus methods for measuring psychoacoustic features [6]. Adaptive up–down methods are able to automatically concentrate trials within dynamic range of psychoacoustic features [7]. Psychoacoustic features in music that affect musical emotions include loudness, pitch level, pitch contour, tempo, texture and sharpness [8].

A variety of experimental designs have been established in the field. In [9-11], tests using a double-blind, double-reference technique were conducted to guide a listener to make consistent judgement on music stimuli. The tests were completed in a number of trials. In each trial, a number of music stimuli were presented in a random order to a listener, who was instructed to compare one or more properties among the stimuli. One of the stimuli was an Explicit Reference Signal (ERS) that a listener was to compare the rest of the stimuli against. The ERS served as a baseline for comparison and was always placed as the first stimulus in presentation. The rest of the stimuli were randomly permutated. The permutation changed from trial to trial. Furthermore, one of the stimuli was a Hidden Reference Signal (HRS), which had well-known music quality. Therefore, each trial had n stimuli, which included one ERS, one HRS at an unknown place, and $n-2$ stimuli of research interest. The listener knew the existence of the ERS, but was not aware of the existence of the HRS. During a test, the listener was asked to compare the quality of $n-1$ stimuli against the ERS, and rank them accordingly. In summary, the ERS differed from the HRS in that: (1) the ERS and HRS had significantly different quality, i.e. if the ERS had a good quality, then the HRS had a bad quality; and (2) the ERS was always the first stimulus and the listener knew of its existence prior to testing, whereas the HRS was randomly placed among the succeeding stimuli, and its placement varied from trial to trial and remained unknown to the listener. The ERS and the HRS served as two references that perceptually calibrated the subjective quality scale of the listener; increasing the stability of the listener's judgement within a trial. It also provided a means of measuring the listener's subjective scale based on the dynamic range that the listener used to rank a well-known good signal vs. a well-known bad signal [9-11].

Studies have found that in normal-hearing listeners, scores in melody tests for the left ear was higher than scores for the right ear [12]. These findings were related to the different roles of the right and left hemispheres of the brain in nonverbal perception. We hypothesized that utilizing two independent equalization curves binaurally may result in improved sound quality.

In this paper, we investigate the effect of personalized dichotic equalization curves on perceived music quality, using adaptive psychoacoustic test methods and a double-blind, double-reference experiment design. We presented listeners with either the same equalization curve in both ears, or two different equalization curves. In the following sections, first we present the experiment methods, including stimuli, contours of equalization curves, in Section 2. Then we propose a double-blind psychoacoustic test framework. Test results are presented and discussed in Section 3, and finally we conclude with Section 4.

2. METHODS

2.1. Stimuli

All testing materials were lossless music sampled at 44.1 kHz. The songs were randomly taken from 70 unprocessed studio master tracks of pop, rock, R&B, Jazz and classical music. From each song, 16 excerpts that each lasted 4~8 seconds long were extracted for the experiment. All music excerpts were pre-processed to maintain the same RMS (root-mean-square) values, and were delivered at 80 dB SPL.

For a music rating test, an excerpt should not exceed 20 seconds to avoid listener fatigue and to reduce the total duration of a test [9, 10]. In this study, all excerpts satisfied this requirement. Additionally, all excerpts met the following two criteria:

- No phrase in the excerpts was interrupted.
- No hearing artifacts (e.g. "clicks") were heard while switching instantaneously among the excerpts.

2.2. Equalization Curves

Digital music equalization was performed in real-time during playback in the frequency domain. Songs were first transformed by Fast Fourier Transform (FFT) into frequency domain, and the output spectrum was the input spectrum multiplied with a specific pair of EQ curves.

A pool of five EQ curves was used in the experiment. As shown in Fig. 1, the five EQ curves were "Bass Boost", "Treble Boost", "Midrange Dip", "Midrange Boost" and "Transparent". The "Transparent" curve had a flat frequency response, which actually did not perform any equalization on an input. For the other four curves, the dynamic range D_{dr} was 9 dB or 4.5 dB.

The "Midrange Dip" was derived from a psychoacoustic-based 80-phon equal loudness contour (ELC) of the ISO-226 standard [4]. An equal-loudness contour is a measure of sound pressure over the frequency spectrum, for which listeners perceive a constant loudness when presented with pure tones. All pure tones on an 80-phon contour $E(f)$ sound equally loud as an 80 dB SPL 1 kHz pure tone. The dynamic range D_{elc} of $E(f)$ is 32.6 dB. Thus, the "Midrange Dip" in Fig. 1 was obtained by:

$$C_{mid-dip}(f) = \frac{D_{dr}}{D_{elc}} \cdot E(f) \quad (1)$$

The "Midrange Boost", also psychoacoustic-based, was derived by reversing the contour of Eq. (1):

$$C_{mid-boost}(f) = -C_{mid-dip}(f) \quad (2)$$

The "Midrange Boost" can boost the mid-range frequencies for listeners who prefer enhancing the vocal component over the instrumental component, if both components exist in music. On the contrary, the "Midrange Dip" enhances the instrumental components.

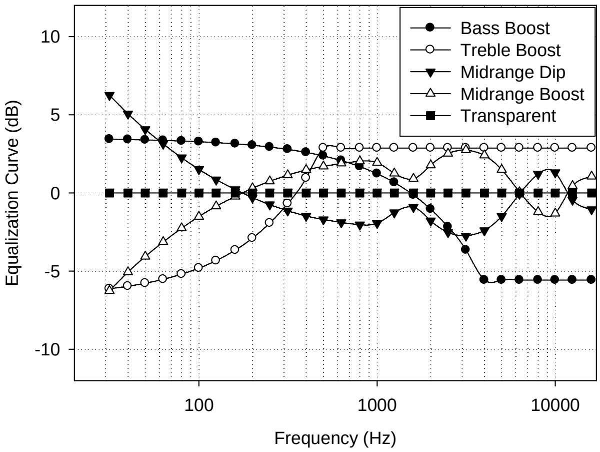

Figure 1: *A Pool of EQ Curves.*

2.3. Subjects

Since non-expert listeners are representative of the general population, subjects both with and without music background were included in the experiment. More than twenty subjects are recommended to reflect qualities of the general population [9, 10]. In this study, thirty-one subjects between the ages of 18 and 43 participated in the experiment. Audiograms of subjects were obtained. All subjects had normal hearing, defined as having pure-tone air conduction thresholds better than or equal to 20 dB HL at octave frequencies from 250 to 8000 Hz [13].

As was established in a variety of studies, both musical experience and musical attitude are critical in identifying personal musicality [14]. Demographic information such as individual musical experience and attitude were collected to investigate the relationship between individual demographics and equalization preferences. Listeners were asked about their musical training and listening habits through two interview questions:

- **Interview Question 1**: *In total, how many years have you spent in (i) playing instruments, (ii) attending music lessons or classes, (iii) participating in musical ensembles (e.g. band, choir or orchestra, etc.), (iv) attending music appreciation classes (e.g. general music classes, music appreciation or theory, etc.), and/or (v) teaching music classes? Please also write a brief description of your musical experience.*
- **Interview Question 2**: *How many hours do you listen to music daily? (a) None, (b) 0.5, (c) 1, (d) 1.5, (e) 2, (f) 2.5, (g) 3, (h) 3.5, (i) 4, (j) 4.5, (k) 5, (l) 5.5, (m) 6, (n) 6.5, (o) 7, (p) 7.5, (q) 8, (r) 8.5, (s) 9, (t) 9.5, (u) 10, (v) 10.5, (w)11, (x) 11.5, (y) 12 or more.*

Question 1 assessed musical experience by asking for the extent of experience in a variety of musical activities. Question 2 evaluated musical attitude by quantifying daily involvement with music appreciation. Among 31 subjects, 16 did not possess any music background, while 15 had music experience ranging from 3 months to 30 years. Detailed demographic information collected from the subjects is shown in Table 1.

Table 1: *Demographic Information.*

ID	Gender	Music Background (Years)	Daily Listening (Hours)
S1	M	None	8.0
S2	M	None	2.0
S3	F	None	4.5
S5	M	None	1.5
S6	M	None	3.0
S8	F	None	2.0
S10	F	None	2.0
S11	F	None	1.0
S12	F	None	4.0
S14	F	None	1.5
S15	M	None	2.5
S17	F	None	1.0
S19	F	None	0.5
S23	F	None	4.0
S29	F	None	0.0
S30	F	None	0.5
S4	F	Played flute (1 year)	5.5
S7	F	Plays instruments including piano, saxophone, clarion, flute, and guitar (12 years)	3.0
S9	F	Had training in piano (4 years)	3.0
S13	F	Performs as a singer and played violin (2 years)	4.5
S16	M	Plays flute (10 years)	3.0
S18	F	Played piano (9 years)	1.0
S20	F	Plays instruments (15 years)	1.0
S21	F	Played piano (3 months)	1.5
S22	F	Plays guitar (3 years)	3.0
S24	F	Plays piano (10 years)	1.5
S25	F	Played flute (5 year)	2.5
S26	F	Plays violin and piano (10 years)	0.5
S27	M	Plays piano (15 years)	2.0
S28	M	Is a piano teacher (30 years)	4.0
S31	F	Plays piano, clarinet and violin (10 years)	2.5

2.4. Experiment Design

Prior to the test, a pool of five EQ curves was established. Each EQ curve had a dynamic range, defined as the difference in decibel (dB) between the highest and lowest magnitudes of the curve. All EQ curves were pre-processed to have the same dynamic range of D_{dr} dB.

Session 1 searched for a pair of best EQ curves (l_f, r_f), where l_f and r_f represent the equalization curves chosen for the left and right ear, respectively. In Session 2, the listener was asked to rate quality of original music and equalized music in a double-blind manner. Therefore, Session 1 was a personalization session, whereas Session 2 was an evaluation session.

In order to find the best EQ curves adaptively, Session 1 was divided into two stages: Coarse Personalization (CP) stage (Stage

1), and fine personalization (FP) stage (Stage 2). In the CP stage, the best pair of curves (l_c, r_c) was chosen from all possible combinations of EQ curves in the pool. In the FP stage, the selected pair of curves was narrowed down to within half dynamic range $\frac{D_{dr}}{2}$ around the pair of EQ curves (l_c, r_c) found in the CP stage. The search result at the end of the FP stage was the final personalized EQ curves (l_f, r_f).

2.4.1. Session 1

The total number of possible combinations of EQ curves was 25, one of which was the combination where the "Transparent" curve was delivered to both ears, which is essentially the same as the ERS. This left a total of 24 possible EQ curves to compare with the ERS. An example combination is, "Bass Boost" delivered to the left ear and "Midrange Dip" delivered to the right ear.

2.4.1.1 Stage 1, CP Stage

Twenty-four combinations of candidate EQ curves were randomly divided into three trials – Trial 1, Trial 2 and Trial 3. Each of the three trials contained a Hidden Reference Signal (HRS), an original excerpt low pass filtered at 3.5 kHz, which served to enhance individual stability. Therefore, each of the three trials had 10 stimuli, which included eight pairs of candidate EQ curves, one HRS, and one ERS. In the CP stage, the pair of EQ curves (l_c, r_c) that provided the highest quality out of 25 possible pairs of EQ curves was determined, where $l_c \in [1, 5]$ and $r_c \in [1, 5]$ represented the indices of EQ curves in the pool for left and right ears, respectively. Music obtained by equalizing with (l_c, r_c) was referred to as coarse personalized music.

2.4.1.2 Stage 1, FP Stage

In Stage 2, Fine Personalization (FP) stage, the pair of EQ curves (l_c, r_c) was refined. The FP stage consisted of only one trial – Trial 4. Assuming the EQ curve l_c had a spectral contour $C(f, l_c)$, a new EQ curve was created as $\frac{C(f, l_c)}{2}$, labelled $\hat{l}_c$, which possessed half the dynamic range $\frac{D_{dr}}{2}$, i.e. 4.5 dB. Compared with the EQ curve found in the CP Stage, this curve offered the same contour, but with narrower dynamic-range. This allowed a subject to fine-tune the search. Similarly, a new EQ curve based on r_c was constructed with half dynamic range $\frac{D_{dr}}{2}$, labelled $\hat{r}_c$. In this trial, in addition to the HRS, a subject compared four pairs of EQ curves – (l_c, r_c), $(\hat{l}_c, r_c)$, $(l_c, \hat{r}_c)$ and $(\hat{l}_c, \hat{r}_c)$ against the ERS. In the FP stage, the best pair of personalized EQ curves was chosen, labelled (l_f, r_f), which was fine-tuned from the coarse personalized curves. Music obtained by equalizing original music with the pair of personalized EQ curves (l_f, r_f) was referred to as fine personalized music.

2.4.2. Session 2

Session 2 was conducted 4 weeks after subjects completed Session 1. The objective in Session 2 was to evaluate the listeners' preferences for equalized music in a double-blind test, and to test the reliability of sound quality rating across experiment sessions.

Session 2 composed of three trials – Trials 5, 6 and 7. In each trial, there were four stimuli, including the Hidden Reference Signal (HRS), the original music, the coarse personalized music, and the fine personalized music. One of the latter three was randomly selected as the "explicit reference signal".

2.5. Experiment Procedure

The experiment consisted of two sessions, Session 1 and Session 2. Both sessions were comprised of multiple trials. In each trial, subjects compared up to 9 stimuli against the ERS. All stimuli in each trial contained the same music, derived from the same excerpt, but filtered by different pairs of EQ curves. The excerpts changed from trial-to-trial and were randomly selected from 70 unprocessed studio master tracks, with no repetitions across trials or across sessions.

The two sessions utilized dichotic presentation, in which two separate and possibly different EQ curves were applied binaurally. Subjects may listen to the ERS and the other stimuli in any order for any number of times, until they were satisfied with the rating. All subjects were instructed to compare music quality based on their own preference. If the music quality of a stimulus was better than the reference, subjects rated it between 0 and 100 by dragging the corresponding slider up to a position that reflected the perceived quality on an ITU-R-BS.1284 rating scale [15]. Otherwise, subjects rated it between -100 and 0 by dragging the slider down. No feedback was provided during testing. The experiment was self-paced, and subjects were encouraged to take breaks at any time.

Because subject fatigue is a factor that can interfere with individual judgments, a music rating session typically should last less than 10~15 trials [9, 15]. In this study, Session 1 included four trials, and Session 2 included three trials, which meet this requirement. Prior to formal tests, preliminary tests showed that Session 1 took about 3~5 minutes, whereas Session 2 took about 2~4 minutes. Both were less than 15~20 minute boundary for listener fatigue [9, 15].

The graphical user interface (GUI) is shown in Fig. 2. There were up to 10 buttons, labeled "*REF*", "*A*", "*B*", …, and "*I*", respectively. Clicking on a button played the corresponding stimulus. The button *REF* contained the ERS, which was the original music in Session 1, and either the original music, the coarse personalized music, or the fine personalized music in Session 2. One of the other buttons contained the HRS, and the rest of the buttons contained an excerpt obtained by equalizing the original excerpt with a specific pair of EQ curves. The "Next" button at the bottom of the GUI allowed the listener to proceed to next trial.

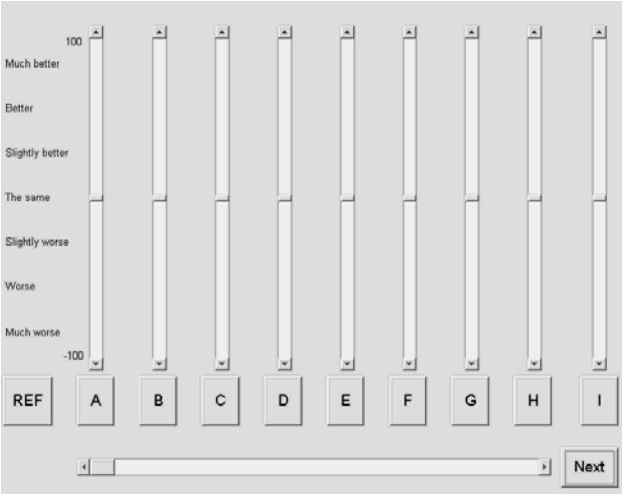

Figure 2: *Graphic user interface (GUI) of a double-blind, double-reference procedure with dichotic presentation.*

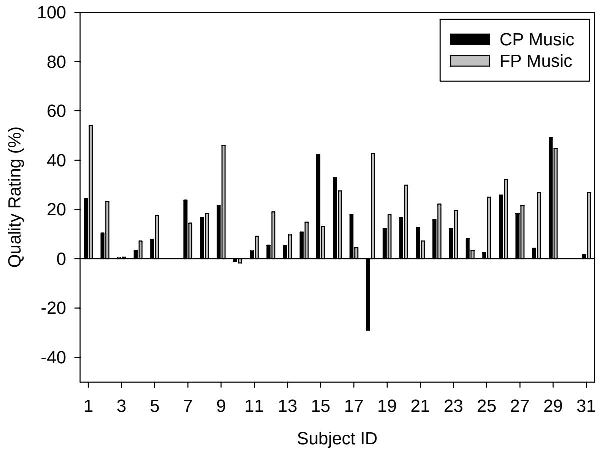

Figure 3: Quality rating of coarse personalized (CP) and fine personalized (FP) music of individual subjects.

2.6. Software and Calibration

The GUI-based program used in the experiment was written in Matlab^{TM}. The program performed equalization in real-time and also recorded subjects' responses. Statistical analysis was carried out in SPSS^{TM}. The music was played from a computer to Sennheiser HDA-200 Circumaural Headphones that were worn by listeners. Prior to the tests, a 1 kHz pure tone was applied to calibrate the software.

3. RESULTS AND DISCUSSIONS

3.1. Quality Rating

Figure 3 shows average quality ratings of CP and FP music with the original music as the baseline for individual subjects. For a particular trial, the stimulus presented as the reference had a relative rating of zero since subjects rated stimuli relative to the reference. If the reference was associated with the original music, the rating for the original music was taken to be zero and the ratings of CP and FP music were taken from listeners' ratings as is. If the reference in a trial was associated with CP or FP music, the rating for the original music, CP and FP music were normalized to zero by subtracting the rating of the original music.

Figure 4 shows average quality ranges of CP and FP music across all subjects. Results indicated that CP music improved the quality rating by 12.1% (SD=0.1443) compared to the original music, and FP music improved the quality rating by 19.3% (SD=0.1440). T-tests showed that these effects were significant ($t(30) = 4.68$, $p < 0.001$ and $t(30) = 7.46$, $p < 0.001$ respectively).

Of the 19.3% in improvement in FP music quality ratings over the original music, 12.1% can be attributed to CP personalization from Stage 1 of Session 1, and 7.2% can be attributed to FP personalization from Stage 2 of Session 1. T-tests confirmed that FP quality ratings were significantly better than CP quality ratings ($t(30) = 2.33$, $p < 0.05$). This suggests that after an initial pair of personalized EQ curves was found, fine-tuning the curve with narrower dynamic range further improved subjective music quality significantly.

A two-way ANOVA was conducted to examine the interaction between gender and degree of personalization (CP or FP) on the quality rating. The quality rating was found to be normally distributed for the groups formed by the combination of gender and degree of personalization as assessed by Shapiro-Wilk test. There was homogeneity of variance as assessed by Levene's test for equality of error variances. There was no significant interaction between the effects of gender and degree of personalization on the quality rating ($F(1, 58) = 0.035$, $p = 0.852$).

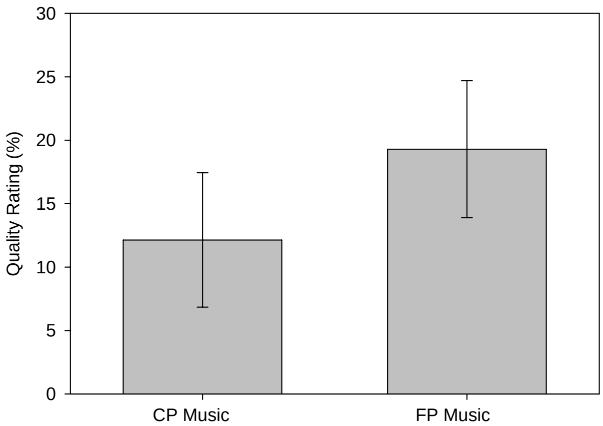

Figure 4: Average quality rating of coarse personalized (CP) and fine personalized (FP) music across thirty-one subjects.

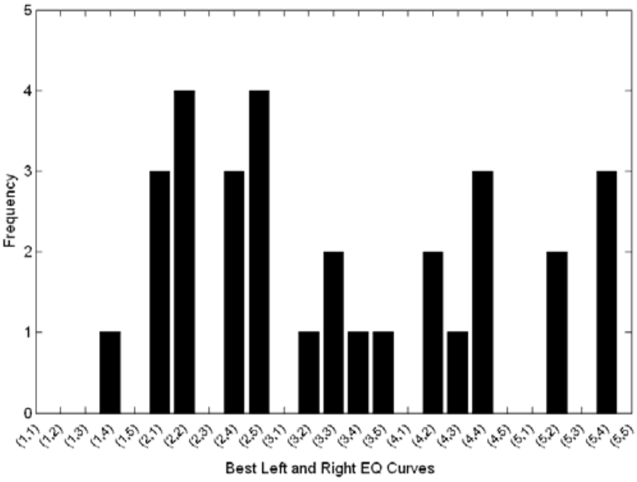

Figure 5: Histogram of best left and right EQ curves.

Since majority of the quality improvement due to personalization can be attributed to Stage 1 of Session 1, it is of interest to examine a histogram of subjects' EQ curve preferences. The histogram is shown in Fig. 5, where the x-axis represent pairs of equalization curves chosen, labelled by a pair of indices (l_c, r_c), where $l_c \in [1, 5]$ and $r_c \in [1, 5]$. The indices 1 to 5 correspond to "Bass Boost", "Treble Boost", "Midrange Dip", "Midrange Boost" and "Transparent" EQ curves, respectively. Fig. 5 shows that none of thirty-one subjects preferred EQ curves to be ("Transparent", "Transparent"). In other words, all subjects preferred equalization over no equalization.

3.2. Predictability of Quality of CP and FP Music

In literature, musical experience and music attitude have been shown to impact individual musicality [14]. We investigated the relationships between musical experience, musical attitude, and the degree of quality improvement obtained for personalized music.

Results showed that neither daily listening time ($r(31)$ = 0.005, ns) nor musical experience ($r(31)$ = 0.330, ns) correlated significantly with the quality rating of FP music. Both daily listening time and musical experience also did not significantly correlate with the quality improvement obtained for personalized music. Now, a question arises: *does the quality rating obtained for personalized music correlate with a joint vector formed by previous musical training and present daily listening time*?

To investigate the question, a joint index s was formed:

$$s = \alpha \mathbf{W}^T \cdot \begin{bmatrix} t_i \\ t_d \\ 1 \end{bmatrix}, \qquad (3)$$

where t_i is the number of musical experience in years, t_d is the number of daily listening time in hours, α is a constant scaling factor, and $\mathbf{W}$ is a weighing vector defined as:

$$\mathbf{W} = \begin{bmatrix} 1 & w_d & \beta \end{bmatrix}^T. \qquad (4)$$

The joint index s represents unique characteristics of individual music experience and attitude.

To explore the question, a joint index $\hat{s} = t_i + t_d$ was first investigated by giving equal weights to training and daily listening, i.e. by assigning $w_d = 1$, $\alpha = 1$ and $\beta = 0$. Correlation analysis revealed that the quality rating for personalized music significantly correlated with $\hat{s}$ ($r(31)$ = 0.374, $p < 0.05$). A higher value of $\hat{s}$ reflected a higher quality improvement brought by the double-blind, double-reference music personalization. Therefore, $\hat{s}$, calculated linearly from individual musical experience and daily listening time, can predict the quality improvement brought by music personalization. In summary, while previous training alone and present daily listening time alone did not correlate with the quality improvement obtainable from music personalization, the combination of both factors did correlate with the quality improvement, and can be used to predict the effect of the music personalization.

In light of the above, regression analysis was conducted to model the dependency of the quality improvement on the weighting vector $\mathbf{W}$ and the scaling factor α, both of which influenced s through Eqs. (3-4). According to the regression analysis, the best-fit model was

$$s^* = \alpha^* \cdot (t_i + 1.125 t_d + \beta^*) \qquad (5)$$

where $\alpha^* = 0.005$ and $\beta^* = 30.75$. Then, correlation analysis was run to investigate the relationship between s^* and the quality improvement brought by personalized music. Results showed a significant correlation between s^* and the ratings given for personalized music ($r(31)$ = 0.380, $p < 0.05$). Therefore, Eq. (5) can be used to predict the quality improvement obtainable from personalized music over the original music for an individual. Specifically, a higher value of the index s^* calculated from Eq. (5) indicates a higher quality improvement.

3.3. Consistency of Ratings of Music Quality

Fig. 6 shows the quality ratings from the CP stage of Session 1 for one particular subject. In Fig. 6, the x-axis shows the index of the selected left EQ curve, the y-axis shows the index of the selected right EQ curve, and the z-axis shows the quality rating. The highest rating found at the end of the CP stage is marked in Fig. 6 by an asterisk. For this subject, the highest rating was located at the pair of EQ curves (2, 5). A fine-tuned search around the pair of EQ curves (2, 5) is conducted during the FP stage of Session 1, leading to fine personalized music for the subject. Fig. 7 shows the quality ratings for FP music from Session 1 for thirty-one subjects.

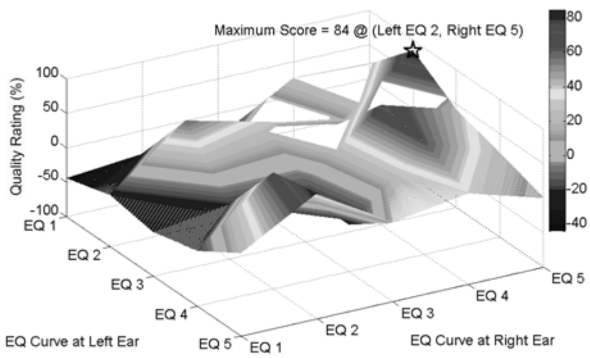

Figure 6: Quality ratings across 25 combinations of left and right EQ curves.

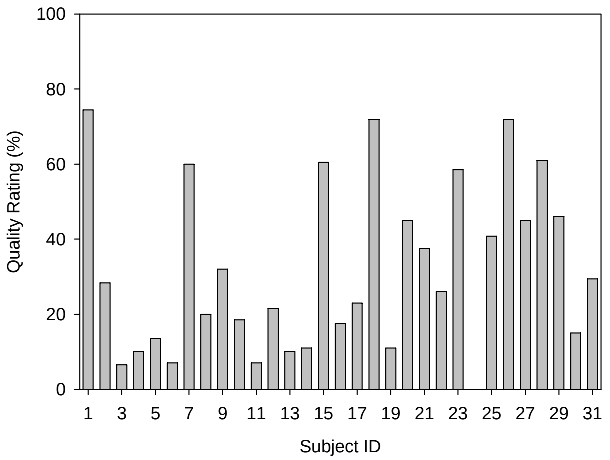

Figure 7: Quality rating for fine personalized (FP) music from Session 1.

To evaluate the consistency of quality ratings for personalized music, ratings of FP music from Session 1 was compared against those from Session 2, which were collected four weeks later. The comparison was analyzed using Cronbach's alpha, a coefficient that summarizes consistency into a number from 0 to 1 [16], with higher Cronbach's alpha showing higher consistency. A Cronbach's alpha value of 0.7 has been established as the usual threshold for consistency in auditory experiments [17, 18]. Cronbach's alpha between the ratings from Session 1 and Session 2 was calculated to be 0.75, which indicated consistency between the ratings collected four weeks apart.

4. CONCLUSIONS

In this study, it was found that personalized equalization (EQ) curves significantly improved the quality of original high-fidelity music by 19.3%. Furthermore, the percentage of quality improvement that a listener can obtain from personalized EQ curves can be predicted from a combination of the listener's musical experience and daily music listening time. Results indicated that listeners in general tend to prefer personalized EQ curves, and tend to be consistent in their preferences for EQ curves.

5. REFERENCES

[1] J. D. Reiss, "Design of Audio Parametric Equalizer Filters Directly in the Digital Domain," *IEEE Trans. Audio Speech Lang Processing*, vol. 19, pp. 1843-1848, Aug. 2011.

[2] S. Cecchi, L. Palestini, E. Moretti, and F. Piazza, "A New Approach to Digital Audio Equalization," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, Oct. 2007, pp. 62-65.

[3] S. M. Kuo and G. H. Canfield, "Dual-channel audio equalization and cross-talk cancellation for 3-D sound reproduction," *IEEE Trans. Consumer Electron.*, vol. 43, pp. 1189-1196, 1997.

[4] S. Bharitkar and C. Kyriakakis, "A comparison between multi-channel audio equalization filters using warping," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, Oct. 2003, pp. 63 - 66.

[5] B. Bank, "Perceptually Motivated Audio Equalization Using Fixed-Pole Parallel Second-Order Filters," *IEEE Signal Processing Letters*, vol. 15, pp. 477 - 480, 2008.

[6] J. C. Baird and E. Noma, *Fundamentals of scaling and Psychophysics*. New York: J. Wiley & Sons, 1978.

[7] H. Dai, "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods," *J. Acoust. Soc. Am.*, vol. 6, pp. 3135-3140, 1995.

[8] E. Coutinho and A. Cangelosi, "Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements," *Emotion*, vol. 11, pp. 921-937, Aug. 2011.

[9] ITU, "ITU-R BS.1116-1 Recommendation - Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," *International Telecommunication Union*, 1997.

[10] ITU, "ITU-R BS.1534-1 Recommendation - Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems," *International Telecommunications Union*, 2003.

[11] ITU, "ITU-R BS.1285 Recommendation - Pre-selection methods for the subjective assessment of small impairments in audio systems," *International Telecommunication Union*, 1999.

[12] D. Kimura, "Left-right differences in the perception of melodies," *Quarterly Journal of Experimental Psychology*, vol. 16, pp. 355-358, 1964.

[13] ANSI, "Specification for Audiometers - S3.6," American National Standards Institute, New York, 2004.

[14] E. T. Gaston, *A Test of musicality*. Lawrence, Kansas: Odell's Instrumental Service, 1958.

[15] ITU, "ITU-R BS.1284-1 Recommendation - General methods for the subjective assessment of sound quality," *International Telecommunication Union*, 1998.

[16] L. J. Cronbach and R. J. Shavelson, "My Current Thoughts on Coefficient Alpha and Successor Procedures," *Educational and Psychological Measurement*, vol. 64, pp. 391-418, 2004.

[17] L. A. Christopherson and L. E. Humes, "Some psychometric properties of the Test of Basic Auditory Capabilities," *Journal of Speech, Language, and Hearing Research*, vol. 35, pp. 929-935, 1992.

[18] G. R. Kidd, C. S. Watson, and B. Gygi, "Individual differences in auditory abilities " *J. Acoust. Soc. Am*, vol. 122, pp. 418-435, 2007.