Audio Engineering Society

Convention Paper 9085

Presented at the 136th Convention
2014 April 26–29 Berlin, Germany

A Loudness-based Adaptive Equalization Technique for Subjectively Improved Sound Reproduction

Konstantinos Drossos[1], Andreas Floros[1], and Nikolaos - Grigorios Kanellopoulos[2]

[1]*Lab of Audiovisual Signal Processing, Dept. of Audiovisual Arts, Ionian University, Corfu - Greece*

[2]*Dept. of Audiovisual Arts, Ionian University, Corfu - Greece*

Correspondence should be addressed to Konstantinos Drossos (`kdrosos@ionio.gr`)

ABSTRACT

Sound equalization is a common approach for objectively or subjectively defining the reproduction level at specific frequency bands. It is also well-known that the human auditory system demonstrates an inner process for sound- weighting. Due to this, the perceived loudness changes with the frequency and the user-defined sound reproduction gain, resulting into a deviation of the intended and the perceived equalization scheme as the sound level changes. In this work we introduce a novel equalization approach that takes into account the above perceptual loudness effect in order to achieve subjectively constant equalization. A series of listening tests shows that the proposed equalization technique is an efficient and listener-preferred alternative for both professional and home audio reproduction applications.

1. INTRODUCTION

Equalization is considered a trivial task for experts and laymen. It is used in almost all audio engineering fields and everyday listening with most of the sound-listeners have somewhat used or it is very likely that they will use an equalizer (EQ) in the future. It is also included in most sound reproduction apparatuses as a part of the electroacoustic reproduction chain. Portable audio playing devices (e.g. iPod or MP3 players), mobile smart-phones, televisions or even car audio reproduction equipment include EQ functionality.

It is well-known that equalization applies an increment (or reduction) of the sound level in specific Frequency Bands (FBs) of the audio content reproduced through an electroacoustic reproduction chain. It is used for adjusting the perceived loudness for each FB available through the EQ settings, following subjective criteria. Nevertheless, upon the variation of the reproduction-chain output gain, the desired perceived loudness for each FB is altering.

The human auditory system incorporates a sound weighting process which is very similar in nature to equalizing. This process is directly related to the weighting of the perceived loudness as a function of frequency and the overall sound reproduction level. It is modeled by the widely-known Equal-Loudness Contours [1]. In practice, the aforementioned human hearing characteristic interferes with the selected EQ settings as the Overall Reproduction Gain (ORG) changes. The variation of the perceived loudness in the lower FBs as a function of ORG is probably the most noticeable effect. Thus, the applied EQ settings will no longer reflect the intentioned analogy of positive or negative gain for the selected FBs.

In this work we present an adaptive equalization approach based on subjective loudness, that aims to compensate the above ear weighting effect in order to retain constant the intended analogy of amplification imposed by the user-defined EQ settings under any sound reproduction gain. We employ the subjective loudness model, expressed in Sone, and transform the loudness level curves for each FB in first degree polynomials. For the needs of a prototype implementation, the presented system regards only the range of $[40, 80]$ *Phon* and 7 FBs. We further perform a series of subjective evaluation tests aiming to investigate the efficiency of the proposed approach for subjectively optimized equalization.

The rest of this work is organized as following: Section 2 provides an overview of existing techniques that investigate or propose equalization processes targeted to subjectively-improved audio playback. Next, in Section 3 an analysis of the proposed system architecture and prototype implementation is provided. The organization of the subjective performance assessment process is included in Section 4 followed by the summary of the obtained results presented in Section 5. Finally, Section 6 concludes the work and points out further enhancements that can be considered for optimizing the implementation efficiency of the proposed technique.

2. RELATED WORK

Although audio equalization is a common task in the generic audio processing field and many studies exist that focus on various equalization application types [3] that are considered significant by the audio engineering community [2], currently one can notice a lack of published works that are concerned with automatic/adaptive and perceptually constant equalization of an audio signal.

In particular, an automatic equalization process for multi-track mixtures has been proposed in a recently published work [2]. The proposed method aims to achieve an equal average perceptual loudness for all channels in the audio multi-track mixture by employing cross-adaptive methods and loudness curves, as they are defined in [1]. Additionally, in [4] another approach for automatic equalization was presented. It was performed with the usage of Artificial Intelligent (AI) and machine learning algorithms and included a pre-requisite of a training stage for the system [2, 4]. Moreover, [5] proposed another method of equalization that employs loudness measures for audio mixing. This method utilizes perceived loudness in order to perform an enhanced audio down-mix or up-mix while preserving the loudness of the initial audio material.

Loudness is also used in hearing aids equalization. For example, in [6] a method is presented for retaining the perceived loudness of audio stimuli, focused on hearing aids. Loudness curves are used for the calculation of the perceived loudness. The latter is altered in order to enhance the perceived audio stimuli and prevent from hearing damages.

Nevertheless, and according to the authors' current knowledge, there is no other published research concerned with an automatic compensation of the human's ear weighting as the overall gain of an electroacoustic reproduction chain alters. Although the aforementioned works are somewhat based on loudness level and perception, no other work known to the authors has considered the use of perceived loudness as defined in [7] for the aforementioned task. Thus, in this paper we present a system for the automatic compensation of the variating weighting introduced by the human ear as the overall gain of the audio reproduction changes. We consider our investigation on the assumption that listeners tend to indicate analogies of perceived loudness when they utilize an EQ and we employ the perceived loudness as a measure used to compensate them.

3. SYSTEM DESCRIPTION

The system exploits the fundamental attribute of subjective loudness where an increment of Sone values directly reflects the increase at the perceived loudness. Thus, a sound with a level of X Sone will sound twice as louder compared to a sound with $X/2$ Sone [7].

In order to make use of Sone values, the actual Sound Pressure Level (SPL) per FB must be known. For the purposes of the current work, the following realistic assumptions have been made:

1. The users of an EQ are likely to indicate (and also perceive) the difference of level in different FB as analogies
2. Slight and iterated differentiations in perceived loudness of the reproduced audio material are recognized as perceived artistic expression (e.g. the artist is playing softer).

Thus, we used the actual SPL for the frequency of 1kHz only and we inferred the analogies for the employed FB in the presented EQ as the sound reproduction gain changes. The actual SPL is measured at the listener's position for all possible combinations of the 1kHz FB EQ settings and ORG values. Such kind of data are likely to be known by individual manufacturers where the technical characteristics of the entire electroacoustic reproduction chain are accurately measured.

The proposed system implementation incorporates a typical seven octave-band equalizer with constant Q factor equal to 1.3333. The corresponding FBs center frequencies are typical octaves (i.e. 125, 250, 500, 1000, 2000, 4000 and 8000 Hz). The gain Gi that corresponds to the i-th FB is estimated every time the listener changes the ORG, based on the output of the proposed human's ear weighting compensation scheme that employs subjective loudness values measured in Sone [7]. This takes into account the target settings of the EQ and the Sone values that correspond to the previously and newly selected ORG value in order to calculate the Gi values as:

$$Gi[n] = Gi[n-1] + 10log_2\left(S[n]\,S[n-1]^{-1}\right) \qquad (1)$$

where S denotes the loudness Sone values and n is the index of the ORG's consecutive alterations performed by the user.

3.1. System Calibration

The calibration of the system was performed by utilizing a band-passed (at 1kHz) gaussian white noise and a sound level meter. The band-pass filter employed had the same characteristics with the one utilized for the actual equalization process at the 1kHz frequency band. The produced sound pressure level (denoted here as $SPL[n]$) was measured at the listener's position for all n ORG values considered in this work. In particular, we employed 11 ORG settings, with values in the range of $[-80, 0]$ dB relative to Full Scale (dBFS). In addition, the root-mean-square (RMS - $R[n]$) value of the band-passed signal and the corresponding Sone value ($S[n]$) were also calculated.

According to [7, 1] and taking into account Eq.(1), both $SPL[n]$ and loudness level $L_L(i)$ values for the i-th FB should be in the range:

$$40 < SPL[n] < 120\,dB-SPL \qquad (2)$$

$$20 < L_L(i) < 80\,Phon \qquad (3)$$

Thus, the minimum and maximum values of the employed reproduction gain were defined as:

$$min\,(SPL[n]) > 40\,dB-SPL \qquad (4)$$

$$max\,(SPL[n]) < 80\,dB-SPL \qquad (5)$$

In order to compensate the difference between the sound pressure and the corresponding loudness values (in $Phon$) for the i-th FB, following Eq.(4) and(5), the reproduction gain was set to $min\,(SPL[n]) \approx 55\,dB-SPL$ and $min\,(SPL[n]) \approx 75\,dB-SPL$.

As it was mentioned previously, for all n possible ORG values, the $RMS\ R[n]$ values were calculated as well. Both $R[n]$ and $SPL[n]$ values represent the calibration data set of the equalization system.

3.2. Human's Ear Weighting Compensation Scheme

The human's ear weighting compensation scheme was activated upon a user-defined ORG change. It takes into account the previous $SPL[n-1]$, $S[n-1]$ and $R[n-1]$ and new $SPL[n]$, $S[n]$ and $R[n]$ values. In particular, for each ORG variation, the Gi values are calculated using the process described by Algorithm 1, under the following assumptions:

1. Loudness level curves can be approximated by first degree polynomials with relative small error

2. The serial filtering procedures that are applied to the digital sound samples by the frequency responses of electroacoustic reproduction chain and room can be considered as linear for each i-th FB, for each listening position and for relatively small variations of the ORG.

Algorithm 1 Human's ear weighting compensation

Get the $R[n-1]$ and $R[n]$ for the current signal
Quantize $R[n-1]$ and $R[n]$ according to calibration values
Get the $SPL[n]$ and $SPL[n-1]$ calibration values based on the current signal's $R[n]$ and $R[n-1]$ values respectively
Calculate the sone ratio from $SPL[n]$ and $SPL[n-1]$
for $i \leftarrow$ All FB except 1kHz **do**
    $Gi_{N-1} \leftarrow$ the previous gain
    Calculate the weighting factor according to sone ratio and ith FB loudness curve
    $Gi_N \leftarrow Gi_{N-1}$+weighting factor
end for

The first assumption is evaluated next in this Section and is combined with the second one that allows the mapping between L_Li and Gi. The second one represents a necessary condition for establishing the evaluation process of the prototype system, provided that the frequency response of the reproduction equipment is not known. In practice, this assumption can be omitted, since the frequency response of the electroacoustic reproduction chain (including DAC, amplifier and loudspeakers or headphones) is known to the manufacturer, while the room frequency response can be measured through existing calibration technologies employed by major sound technology manufacturers, e.g. [8].

The level of the perceived loudness can be calculated from the Equations provided in [1]. The corresponding values in *Phon* that lay in the range [20, 80] and for all FBs are portrayed in Figure 1. Clearly, all curves in Figure 1 and for *Phon* values in the range [40, 80] seem to be rather straight. Thus, they can be approximated by first degree polynomials, i.e. $f(x) = ax + b$. The a and b factors along with the root mean square errors (RMSE) for such approximation and for all curves in Figure 1 are shown in Table 1. Thus, L_Li can be evaluated from the the sound pressure level in the corresponding FB as:

$$L_Li = aSPLi + b \tag{6}$$

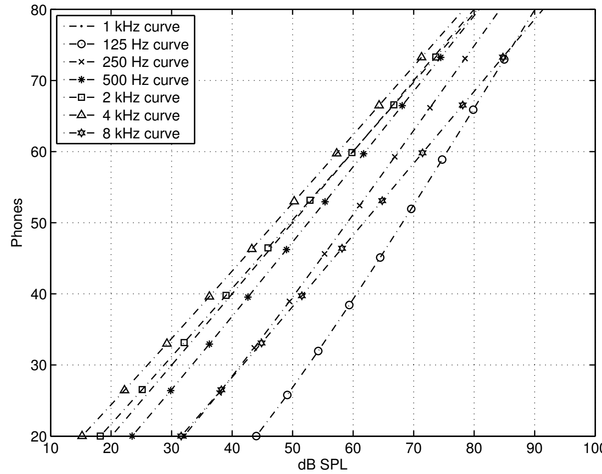

Fig. 1: Phon Vs dB-SPL values, according to [1], for all FBs considered

Table 1: Mean square errors for first degree polynomial approximation for all curves of Fig. 1 and for sound pressure level in the range of [40, 80] dB-SPL.

Freq. (Hz)	RMSE	*a* factor	*b* factor
125	0.09564	1.365	-43
250	0.06311	1.186	-19.97
500	0.02696	1.061	-05.765
1000	0.003387	1	-00.01427
2000	0.0129	0.9688	$+01.948$
4000	0.0228	0.9624	$+04.642$
8000	0.0162	1.012	-12.47

with the a and b parameter values defined in Table 1 for each of the i-th FBs.

Moreover, Gi values (see Eq.(1)) are defined through the actual i-th FB EQ slider positions. But since $S[n]$ values in Eq.(1) are calculated for the 1kHz FB, the direct application of Eq. 1 would result in simply offsetting the same FB EQ settings analogy in proportion to ORG changes. According to assumption 1 and since a for the 1kHz FB is almost equal to 1, Eq.(1) can be written as:

$$Gi[n] = Gi[n-1] + a_i 10 log_2 \left(S[n] S[n-1]^{-1} \right) \tag{7}$$

Clearly, the new Gi value will be based on the previous one plus a weighted factor that is based on the perceived loudness. Eq.(7) was used to calculate the new

FB EQ settings for the *i*-th band, upon an ORG user-defined change.

4. LISTENING TESTS

The performance of the presented system was assessed through a series of listening tests. In an attempt to emulate the typical audio reproduction conditions met in a usual listening environment, the tests were organized in a small office environment. Moreover, no special treatment for the acoustic properties of the room had been made. A total of 8 human subjects participated in the listening tests. Half of them were experienced audio recording, mixing and mastering professionals (or semi-professionals), while the other half were adventitious music listeners. The following subsections describe in detail the overall experimental setup.

4.1. Experimental Setup

The experimental measurements were performed using a common personal computer as a digital audio workstation, equipped with a digital audio interface. Audio playback was performed through a set of active stereo loudspeakers. The listeners were placed on the ideal sweet-spot area imposed by the above stereo setup. A sound level meter was used for performing the system calibration described previously. The selection of the latter as a medium accuracy measurement device was intentional, aiming to achieve a measurement performance equivalent to the one exhibited by level meters incorporated in modern consumer audio equipment. The exact models of the above equipment together with their detailed technical specifications are presented in Table 2.

Three Compact-Disk quality musical excerpts from different musical genres and with a duration in the range of 3 minutes were selected (see Table 3). These were reproduced sequentially in the order appeared in this Table. The above musical excerpts were selected from Discogs [9].

4.2. Experimental Procedure

The experimental process was organized in two steps. During the first one, each participant had to define a preferred equalization scheme for each of the selected musical excerpts at a predefined ORG (which was the same for all test cases and equal to -6.02dB-FS). Next, the system varied the ORG, by randomly selecting it's consecutive values between the corresponding preselected set presented in Table 4. This random selection process included the original -6.02dB-FS ORG setting used for defining the preferred equalization scheme in step one.

Table 2: Technical specifications of the employed equipment

Manufacturer	Model	Tech specs
Lexicon	Omega	Output: 1/4“ TRS balanced outputs, +19 dBu maximum output, D/A: 24bit, 109dB typical, A-weighted, 20Hz - 20kHz.
ESI	nEar 06	Input sensibility: 270mV, Frequency response: 55Hz - 20kHz, Crossover frequency: 2.7kHz
Lutron	SL-4010	IEC 61672 class 2, Resolution: 0.1dB, Measurement Frequency: 31.5Hz - 8kHz

Table 3: Details of the utilized audio material

Genre / Style	Artist	Title
Jazz	D. Ellignton, C. Mingus, M. Roach	Caravan
Rock	The Beatles	Come Together
Trip - Hop	Massive Attack	Safe from Harm

As it will be explained next, this is an important aspect of the experimental process used for verifying that each subject had a clear perceived impression of his/her original equalization preference. Despite the above random selection, all ten possible ORG values within the previous range were applied for each musical excerpt, resulting into a total of 10 experimental repetitions per excerpt (and thus a total of 30 repetitions per participant).

After an ORG value random variation. the proposed human's ear weighting compensation scheme was automatically applied and the participants were asked to manually adjust (if required and according to their perception) the gain setting per frequency band in an attempt to perceptually reattain the original equalization target scheme. According to the above experimental process, if the listening perception achieved by the the applied weighting

Table 4: ORG values indices and their corresponding dB-FS

ORG index	dB-FS
1	-20.00
2	-13.98
3	-10.46
4	-7.96
5	-6.02
6	-4.44
7	-3.10
8	-1.94
9	-0.91
10	0

compensation scheme was close to the original-target EQ settings, only small (or no changes) would be further implied by the human listeners. These frequency-dependent gain changes were tracked and stored for obtaining the results presented in Section 5. Clearly, any changes applied during the -6.02dB-FS case would imply the fact that the specific listener has no clear perceptual feeling of his original equalization setting for the particular reproduced audio track.

The overall experimental process was controlled through a software application developed for the purposes of this work. This application realized all the necessary signal processing algorithms required for equalizing, as well as the user interface for communicating with the participants using the widely used slider-based equalizer control approach. Fig. 2 shows a typical screenshot of this application. Prior to any experimental session, the functional details of the above application were analytically described to each participant.

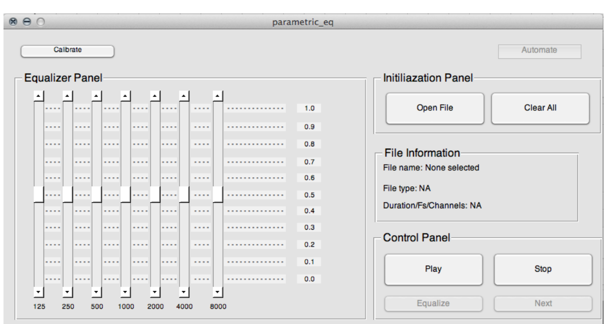

Fig. 2: A screenshot of the application

5. RESULTS

From the complete set of the FB gain changes implied by all participants we measured the deviations from the corresponding values proposed by the system. These deviations can be considered as the perceptual error imposed by the proposed technique. They were expressed in terms of RMSEs. More specifically, the above deviations measurements regarded:

1. The RMSE value for each FB, all musical excerpts and averaged on all ORG values
2. The RMSE per ORG value for each FB and for all musical excerpts

In particular, each EQ frequency band was indexed by an integer $x = 1, 2, \ldots, 7$. For the n-th ORG value alternation, a vector SE_n was defined containing the gain per frequency band values BG_{Sx} as they were defined by the proposed adaptive equalization technique, that is:

$$SE_n := [BG_{S1}\ BG_{S2}\ \cdots\ BG_{S7}] \tag{8}$$

Accordingly, an additional vector UE_n was formed, containing the corresponding gain values BG_{Ux} that were implied by the user corrections:

$$UE_n := [BG_{U1}\ BG_{U2}\ \cdots\ BG_{U7}] \tag{9}$$

Finally, using the vectors SE_n and UE_n a deviation vector was calculated as:

$$D_n = SE_n - UE_n \tag{10}$$

Vector D_n was used for the calculation of the RMSE for the n-th ORG variation session. Table 5 summarizes the RMSEs values averaged for all ORG variations, by considering a) all participants and b) the subjects that were experienced audio engineering experts only. Clearly, all values (except one) are below 1. Since a deviation value equal to 1 corresponds to an one step change of the FB EQ setting, this is a strong indication that the participants' perceptual corrections were minimal and that, consequently, the proposed system adaptation mechanism performed very efficiently in the perceptual domain.

Table 5: RMSE value for each FB, all musical excerpts, averaged on all ORG values and for all participants (A) and experts only (E)

	Average RMSE	
FB (Hz)	**A**	**E**
125	0.9583	0.6655
250	1.1333	0.1366
500	0.9711	0.1852
1000	0.9972	0.9023
2000	0.4798	0.1852
4000	0.9196	0.1631
8000	0.7879	0.4984

The above trend is clearly further improved if only the expert subjects are considered. In this case, a slight increment of the error values can be also observed at the lowest and highest FBs. A similar tendency can be also inferred for the RMSE values from Table 1. These two similar in nature increments in the RMSE values are likely to be well connected, since the *a* and *b* factors in Table 1 are utilized for the curves employed in the human's ear compensation scheme.

Table 6 demonstrates the RMSE values distribution for all ORG values and FBs as an average for the three musical excerpts considered. The results are again organized including a) all participants and b) the audio engineering experts only. The ORG indices presented are previously defined in Table 4.

In the all participants case, it can be observed that there is an increment of the error value while the ORG is reduced. In most frequency bands, greater error values are appeared for the two lower ORG settings. However, this fact is not valid for the expert-participants only, since in this case there are FB and ORG combinations where no human corrections were made to the EQ settings proposed by the human's ear weighting compensation scheme. It should be also noted that the greatest error value was obtained in the 1kHz FB and for ORG=-3.10dB-FS (and from corrections made by experts).

In any case, greater error values are presented in the non-expert subjects. This is also verified for ORG=-6.02dB-FS, the original gain setting that was employed for defining the initial target equalization settings per participant and music track. This clearly indicates that, in general, non-expert participants corrected their own preferences when the latter were presented to them randomly. However, this is not the case for the audio-engineering experts, since the corresponding greater RMSE value for ORG=-6.02dB-FS equals to 0.5.

6. CONCLUSIONS & FUTURE WORK

In this work a system for loudness-based adaptive equalization is introduced, aiming to compensate perceptual variations on the equalization outcome imposed by alternations of the overall reproduction gain. The proposed technique is based on the perceived loudness model and employs a novel scheme for human's ear weighting compensation for exploiting the fundamental attribute of subjective loudness where an increment of Sone values directly reflects the increase at the perceived loudness. Within the framework of this work, a prototype system for realizing the above technique was implemented, incorporating a typical seven octave-band filter-bank.

The performance of the adaptive equalization method was assessed through a sequence of listening tests in which both audio-engineering experts and non-experts participated. During these tests, the participants evaluated the perceptual accuracy of the developed system by applying gain-corrections per frequency band on the corresponding settings proposed by the aforementioned system, for a wide range of overall reproduction gain values.

The results obtained showed that in general, the system perceptual performance is high, since the subjects' changes on the FB gain values were rather limited in number and amount for all ORG values considered. In particular, the audio-experts tend to mostly accept the system's proposed settings compared to non-expert subjects. In addition, non-experts seem to be not comfortable with the ORG alterations as they portray corrections even to the settings made by themselves when the latter are presented randomly in the series of the subjective tests. However, the extent of the above trends is rather limited, rendering the proposed technique an efficient add-on for perceptually optimized equalized audio reproduction.

Further enhancements of the system implementation can be considered as future work, aiming to optimize the levels of the system's efficiency. For example, the number of the frequency bands defined for the equalization process can be increased by considering 1/3 octave frequency bands. It is in the authors' near future intentions to investigate any potential advantages of such extensions. Furthermore, the integration of the proposed sys-

Table 6: RMSE per ORG index (OI) and FB for all musical excerpts and for all participants (A) and experts only (E)

OI	FB center frequency (Hz)													
	125		**250**		**500**		**1000**		**2000**		**4000**		**8000**	
	A	E	A	E	A	E	A	E	A	E	A	E	A	E
1	1.66	0.71	1.44	0.00	1.47	0.29	1.55	1.19	0.80	0.00	1.04	0.00	0.76	0.29
2	0.96	0.29	0.91	0.00	0.76	0.00	1.38	1.26	0.54	0.29	1.77	0.00	1.15	0.29
3	1.00	0.76	1.27	0.00	1.14	0.00	1.06	0.96	0.61	0.58	0.87	0.00	0.76	0.41
4	0.91	0.65	1.21	0.00	1.19	0.41	0.74	0.87	0.79	0.00	0.35	0.29	0.79	1.04
5	0.71	0.50	0.82	0.00	0.79	0.00	0.54	0.00	0.41	0.29	0.87	0.00	0.71	0.29
6	0.58	0.50	1.19	0.29	0.89	0.29	0.58	0.58	0.29	0.29	0.61	0.00	0.84	0.71
7	1.17	0.50	1.40	0.29	1.27	0.29	1.68	2.28	0.35	0.00	1.14	0.00	0.89	0.41
8	1.10	1.39	0.91	0.50	0.89	0.29	1.00	1.32	0.46	0.49	1.00	0.64	0.79	0.64
9	0.70	0.96	1.15	0.29	0.54	0.29	0.79	0.29	0.35	0.00	0.82	0.29	0.68	0.50
10	0.79	0.41	1.02	0.00	0.76	0.00	0.64	0.29	0.20	0.00	0.74	0.41	0.50	0.41

tem functionality (including the initial calibration stage) within existing consumer equipment is a necessary next step that will probably impose specific declinations from the original design, the effect of which should be further exploited and assessed.

7. REFERENCES

[1] Acoustics – Normal equal-loudness-level contours, ISO 226, 2003

[2] P. Gonzalez and J. Reiss, "Automatic equalization of multi-channel audio using cross-adaptive methods", presented at the AES 127th convention, New York, U.S.A., 2009 October 9–11.

[3] J. Mourjopoulos, "Digital equalization of room acoustics", Journal of the Audio Engineering Society, Vol. 42, No. 11, pp. 884-900, 1994.

[4] D. Reed, "A perceptual assistant to do sound equalization", presented at the 5th international conference on Intelligent user interfaces, New Orleans, United States, 2000 January 9–12.

[5] F. Baumgarte, "Equalization for audio mixing", European Patent EP20030253971, Feb. 01, 2004.

[6] D. Kwon, "Equalization method using equal loudness curve based on the IS0266:2003 standard, and sound output apparatus using the same", U.S. Patent 8135146 B2, Mar. 13, 2012.

[7] S. Stevens, "A scale for the measurement of the psychological magnitude: loudness", Psychological Review **43** (1936), no. 5, 405–416.

[8] HARMAN (2014), Premium audio entertainment [Online]. Available: http://www.harman.com/EN-US/Pages/Home.aspx (URL)

[9] Discogs (2014), Database and Marketplace for Music on Vinyl, CD, Cassette and More [Online]. Available: http://www.discogs.com (URL)