Study of Audio Loudness Range for Consumers in Various Listening Modes and Ambient Noise Levels[1]

John Kean*, *AES Member*, Eli Johnson*, *AES Member*, Dr. Ellyn Sheffield**

**NPR Labs, Washington DC;* ***Towson University, Towson MD

Various systems have been developed for consumer audio playback to control loudness range or limit maximum loudness; these systems have generally been ad hoc developments that vary in operational controls and performance. Many audio control systems are designed to focus on the signals, such as electrical signal peaks, rather than effect of loudness variation on listeners. This study investigates the range of audio program loudness that is desired by listeners in typical noise environments. Listening conditions include the ambient noise environments typically encountered by listeners with fixed and portable audio playback equipment, such as average homes, business offices, automobiles and public transit. The results may inform the design of audio loudness management control systems for consumer audio playback systems.

1 INTRODUCTION

This study was conducted to determine the maximum variation of audio loudness or "loudness range" desired by listeners in various listening environments. Techniques to control audio level, such as automatic gain controls, volume limiters, etc., have been introduced into automobiles, computer players and other devices, but relatively little data exists on the audio loudness ranges preferred by listeners in these consumer audio playback systems. This study was intended to provide information to guide the design of loudness management techniques for audio playback systems and headphones for mobile playback devices. The information also may provide information for producers of audio content on how the loudness range of their program material meets the needs of listeners.

At the production and transmission stages of audio content, systems have been available for more than 70 years to control audio signals. For good reason, these systems are focused on the technical needs of the channel in order to avoid overload and to optimize the dynamic range of the medium. Their designs are seldom explored in an empirical way, and have generally been estimates of listener's needs for loudness range, without awareness of the end listening conditions. [1] However, it is likely that the loudness range desired by listeners is dependent on the amount of ambient (background) noise in the listening environment. Ambient noise conditions typically encountered by listeners may include average home noises, business office machines and background chatter, automobile cabin noise and, particular to portable audio equipment, public transit. To investigate the relation between ambient noise and loudness range, this study measured the sound levels desired by listeners in specific noise environments at known sound pressure levels. The audio content was derived from popular music, fine arts performances, newscasts and talk shows currently distributed by Internet audio streams, podcasts, etc. Both loudspeaker and earbuds listening modes were tested.

2 THE EFFECTS OF EXCESSIVE LOUDNESS RANGE

Figure 1 illustrates a condition in which the audio level (variation in loudness) drops below the ambient noise level in some sections. The sample audio was taken from an Internet music streaming service, using a "movie music" genre selection, over a 30-minute period. Both loudness, based on the ITU-R BS.1770 meter, and signal peaks are shown. [2] For the sake of illustration, the change in loudness in this sample is assumed to be greater than listeners prefer. (This is not intended to suggest that listeners should dislike a larger loudness range, which can impart naturalness and even excitement to the sound, in a

[1] Sponsored by the Consumer Electronics Association and reviewed by the CEA's R3 Audio Systems Committee

suitable listening environment. As discussed further below, however, listeners may dislike unnatural or unexpected variations in loudness and alternatively, reaching for the volume control to manage undesired variations in loudness.)

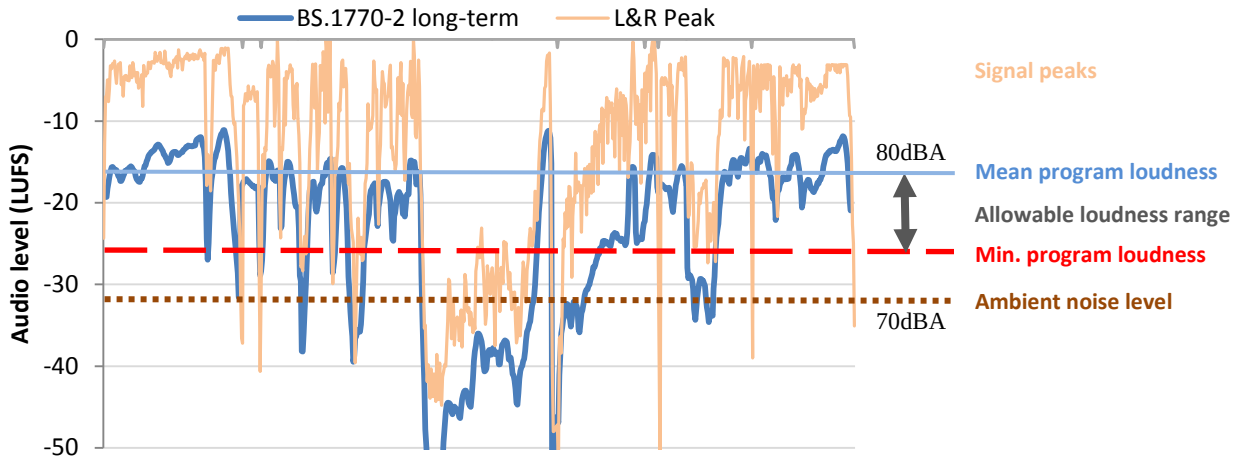


Figure 1 - Measurement of musical content from and online music service over a half-hour period. The sample includes two commercial announcement breaks, indicated by the closely-spaced tick marks on the top line.

The audio content is shown above with the blue horizontal line, measured by the ITU loudness meter using the scale to the left. The mean (integrated) loudness of the entire audio sample is -16 LUFS, shown by the light blue line. Using the sound pressure level scale in Figure 2, "Stereo music" is indicated at a sound pressure range of 60-80 dBA. A background noise environment of 45-67 dBA represents a typical office (contributed by air handling systems, distant conversations, computer equipment, etc.), and so on.

If noise in a car averaged 70 dBA, it is apparent that some sections of the audio would be equal to, or less than the ambient noise. From a review of the SPLs in Figure 2, we can see that some common environments produce sound levels that may prevent some of wide loudness range audio content from being heard.

A related issue with wide loudness range is reducing variations that exceed the requirements of the listener, when program audio with low-level dialog is interspersed with loud dialog or sound effects. Sudden increases may disturb a sleeping partner or neighbor, or simply exceed the tastes of the listener. This can affect the listener's enjoyment, and possibly even render sections of the audio inaudible. Thus, it is important to reduce variations that exceed the requirements of the listener.

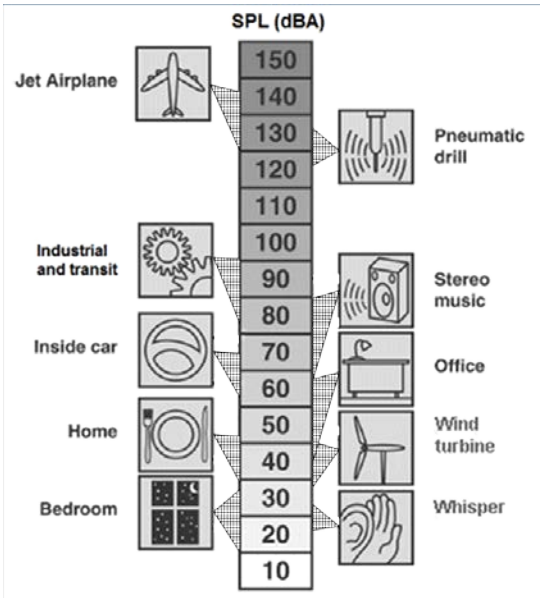


Figure 2 –Some examples of noise environments and ranges of sound pressure level, in dBA.

3 LISTENER CONTROL OF PLAYBACK

Listeners are, of course, at liberty to adjust playback volume, but the necessity to readjust the volume during a program has a cost in terms of annoyance. The authors conducted an earlier test of listener reaction to volume changes during audio streams. [3] The study found the reactions to adjusting the

volume are indicative of how listeners may feel about "gain riding" a program with too much variation in loudness. (Volume changes are not the same as loudness range, as loudness range is usually a natural part of the audio content, while loudness shifts are inadvertent changes in level during a program.)

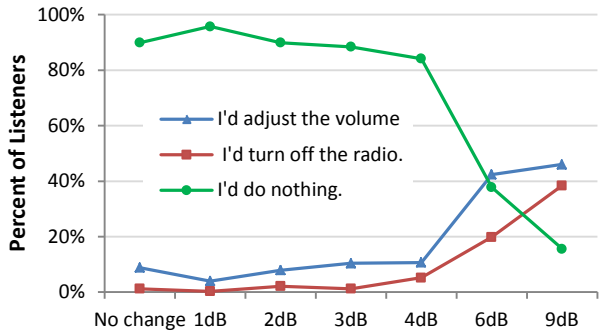


Figure 3 - Listener's behavior to change in loudness IF the change was frequent. Loudness units (LU) with the ITU loudness meter are equivalent to dB

Figure 3 shows that listeners were inclined to "do nothing" until changes in loudness reached 4 LU, above which listeners decided they would adjust the volume knob. At a loudness change of 6 LU, another 20 percent of listeners would "turn off the radio". At a change of 9 LU, the percentage of listeners who would adjust the volume did not increase much, but nearly 40 percent of listeners said they would stop listening. This result suggests that listeners tolerate a degree of loudness range variability, beyond which they may give up on the program.

Figure 4 is a histogram based on the audio levels measured in Figure 1 for loudness and signal peaks. Despite the relatively large loudness range of this audio, the instantaneous signal peaks show a common effect with program audio that has been pre-processed: the distribution of audio peaks end abruptly at some level (near 0 dBFS, in this case) with the majority of occurrences of the signal bunched close to the maximum. The auditory system hears a much more time-averaged signal, with emphasis on frequencies above 2 kHz, as measured by psychoacoustic studies. [4] The ITU loudness measurement uses a frequency weighting filter and time integration to represent the perceived loudness of program audio. In this study, audio loudness was measured and displayed using a Short Term integration time, having a 3-second averaging interval.

Because audio loudness peaks are often distributed around a maximum, the test planning should recognize that listeners usually hear drops below an average level. Realistic audio samples, then, would have decreases below the average, rather than increases above the average.

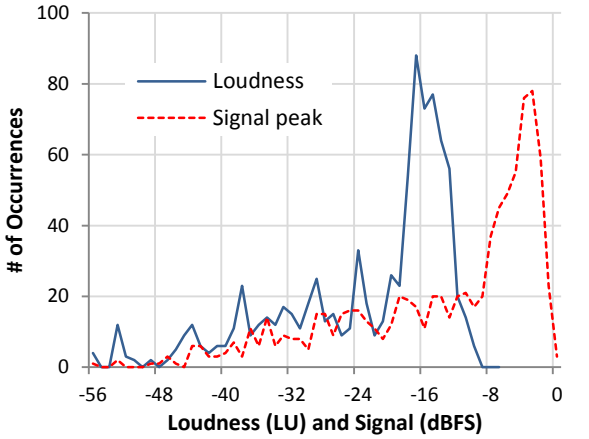


Figure 4 - Histogram of audio loudness from audio measured in Figure 1

4 TEST LAYOUT

The current test presented listeners with a series of one-half to two minute audio samples. Each sample was played back at program levels that vary according to the original production. Listeners were surrounded by noise fields, such as an office environment, roadway noise within a car, etc. These environmental noises may interfere with the subject's hearing of the audio samples. The test subjects were able to readjust playback gain during each sample as necessary to overcome the environmental noise and to suit their taste. Different audio samples were played with different levels of environmental noise during the course of the test. A MATLAB program managed the test playback and continuously recorded the listener's playback settings during each audio sample.

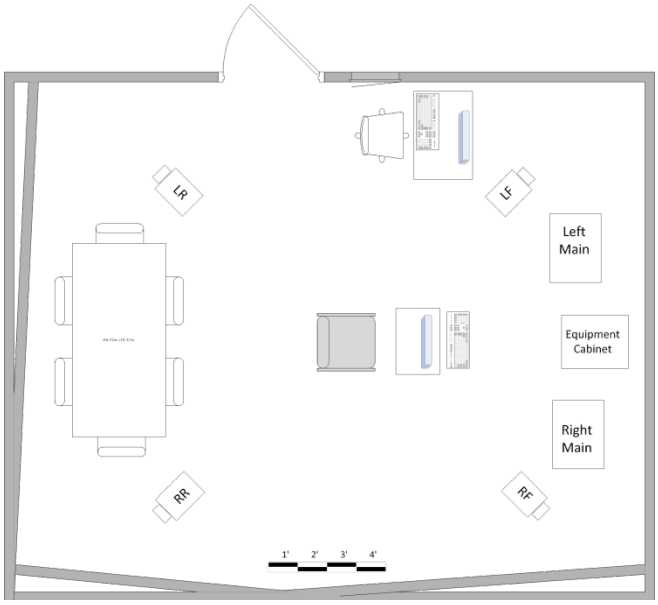


Figure 5 - Layout of audio test room

The test included 37 participants under the age of 65, recruited from the general public and NPR staff. (Those recruited from NPR staff could not be involved with audio engineering as audio engineers may have predispositions to loudness range that may not represent the general public.) The test required less than two hours of participants' time, including instructions by the experimenter and two listening sessions.

Testing was conducted in NPR Labs' Audio Laboratory, an acoustically isolated and treated room approximately 18 feet by 23 feet in size. The layout of the room, shown in Figure 5, placed the test subject in the approximate center of the room sitting in a swivel armchair. The experimenter was positioned to the participant's far left with an equipment cart containing the computer that runs the audio testing and records the test subject's responses.

The test subject was flanked by four Mackie HR824 bi-amplified loudspeakers, labeled "LF", "RF", "LR" and "RR", mounted on stands approximately 3 feet above the floor. The speakers reproduced the environmental noise surrounding the test subject. A low table with a small flat-screen monitor was positioned in front of the listener to view the test questions and control the playback volume. Figure 6 provides a view of the test room from behind the listener.



Figure 6 - Audio test room showing mid-field monitors (yellow cones) and playback volume control (inset)

Half of the audio material during the listening test was heard over loudspeaker. For these tests, the left and right main speakers were KRK Rokit PR10-3 mid-field monitors, a 3-way design with a rated frequency response of 31 Hz to 20 kHz.

The other half of the listening testwas conducted with earbuds, which are popular with users of smart phones, tablets and music players. The device, shown in Figure 7, was the Apple "EarPod®", which was selected because it is one of the earbud products most widely used by consumers, and because it is an open-air type, that is, having minimal isolation (acoustic attenuation) of outside sounds. One study reports the isolation (attenuation) of outside sounds to be less than 2 dB. [5] Noise-blocking earbuds and super-aural (over-the-ear) headphones were unsuitable because they would attenuate the ambient noise environment heard by listeners. This would have varied the results according to the isolation performance of each model of earbuds.



Figure 7 - Apple EarPods

5 LISTENER CONTROLS AND DATA COLLECTION

A small flat-screen monitor was positioned on a small table in front of the listener. On the screen, listeners were able to view on-going test instructions and track their progress through the overall test. Each audio sample was presented in sequence and played by pointing and clicking on a box next to the track label on the screen.

During audio sample playback, listeners had the option to change playback gain of the loudspeaker or earbuds in response to changes they heard in the audio loudness. Their changes were stored by the computer and later compared to the loudness of the sample.

Listeners' playback gain could be controlled by a USB control "knob" on the table in front of them, seen in the inset of Figure 6. Listeners could raise or lower their volume by spinning the knob with their fingers: clockwise to raise loudness and counter-clockwise to reduce loudness, at any time during the sample playbacks. The experimenter instructed each listener on the operation of the screen and playback gain control before their test, and asked them to adjust volume for the most pleasing playback effect. They were informed that they could change their playback level at any time during each sample, but they could leave the level in place if they were satisfied with the sound.

6 NOISE ENVIRONMENT AND PLAYBACK SPL

The A-weighted sound pressure level of the environmental noise, at the position of the listener's head, was measured and set in advance of the experiment. During the test, the environmental noise started at least 20 seconds before the playback of any test material to allow the test subject to become accustomed to the conditions. A Tenma Digital Sound Level Meter, model 72-947, was used with the loudspeaker playback system at the same position of the listeners' head. The SPL of the noise environments, using slow averaging, are listed in Table 1.

Table 1 – Sound pressure levels associated with each noise environment

Noise Environment with loudspeakers	dBA	Noise Environments with earbuds	dBA
quiet office	42	quiet office	42
restaurant chatter	53	restaurant chatter	53
outdoor roadway	65	outdoor roadway	65
vehicle cabin	67	public transit	75

The loudspeaker tests included an automotive environment simulating the in-cabin noise of a vehicle at highway speeds.[2] The quiet office and restaurant recordings were selected from sound effect tracks with care to avoid any distracting or sudden noises that might disrupt the listener's hearing of the program

[2] The stereo recording was taken from an actual passenger car on a concrete surface at 60 MPH. The noise level at the center of the cabin was 67 dBA; the same level was used for playback in the audio environment.

audio. To minimize phase effects from the four loudspeakers producing the environmental noise, the stereo recordings were played back with a small amount of DSP reverberation. The slight reverberation avoided "phasiness" in the sound field when one moved their head around the middle of the sound field.

In lieu of the in-vehicle noise, the ear bud tests included a public transit environment, similar to riding on a modern light rail system. After the tests, however, examination of the data revealed a number of cases where listeners turned up the playback gain to maximum. A smaller number of such cases occurred with the outdoor roadway environment. This did not occur with the loudspeaker tests because higher potential SPL was available. Because it could not be determined if listeners desired more playback gain, a decision was made to exclude all of the public transit and outdoor roadway noise test data from earbuds.[3]

Measurement of the sound pressure level of the listener's loudspeaker playback system was made with one-octave pink noise centered at 1 kHz. As the loudness of the listener's program material was aligned with an ITU loudness meter, the playback SPL was determined with the noise at the target level of -23 LUFS. With the playback gain control at full, the SPL was 81 dBA. For the Apple EarPods, the electrical voltage at the terminals was measured using a 1 kHz tone. With the playback gain control at full, the voltage corresponded to a measured SPL of 71 dB. [6]

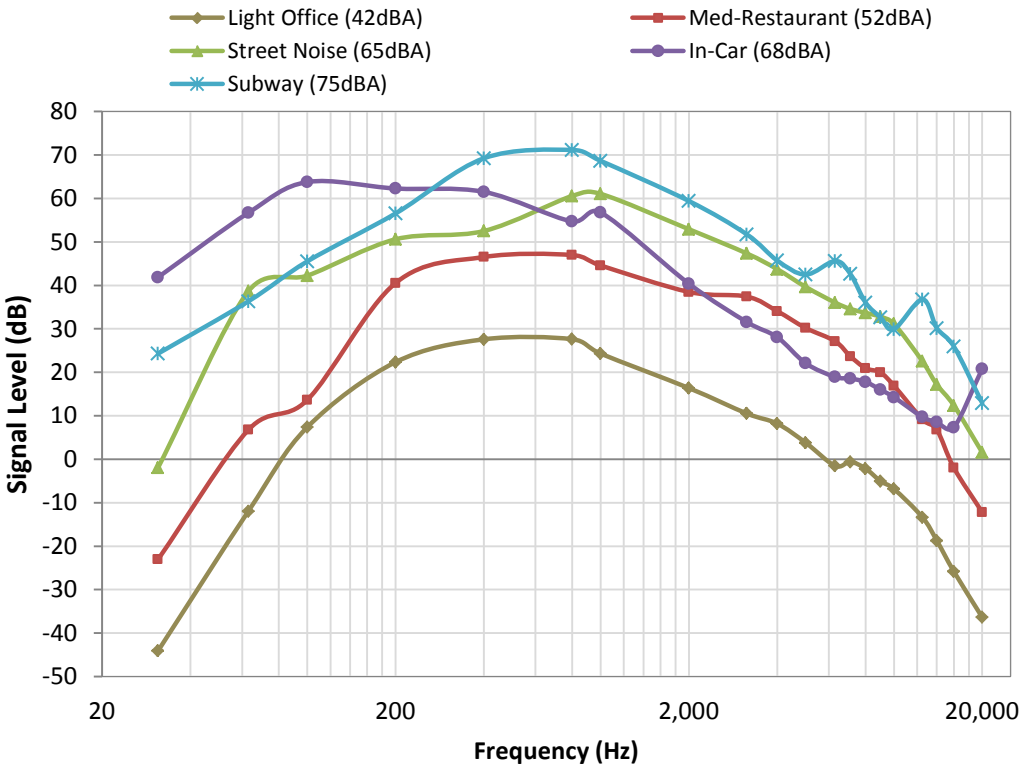


Figure 8 - Spectral distributions of noise environments

The spectral distribution of the noise environments can affect the listener's playback volume preferences, in addition to the single-valued A-weighting for sound pressure level. To help illustrate the differences in

[3] The Apple Ear Pods were driven from the headphone output of a Dell OptiPlex 760 desktop computer running Windows XP. It appears that the audio engine in the computer, in conjunction with the MATLAB test program, had a gain structure such that a -23 LUFS signal produced an output that did not satisfy listeners from time to time.

noise environments, Figure 8 shows the spectral distribution for each of the noise environments.[4] The distribution of the medium-level restaurant environment is similar to the quiet office from 200 to 2,000 Hz, neglecting the 10 dB overall difference in SPL, with a slight rise above 2,000 Hz. The in-car environment, while 16 dB louder than the restaurant environment in A-weighted SPL terms, has the same or less power at frequencies from 2,200 Hz to 7,500 Hz, but nearly 50 dB more power at 40 and 60 Hz. The distributions of street noise and subway (public transit) are relatively stronger in the mid-frequencies from 300 or 400 Hz to 2000 Hz than the light office and restaurant noise.

7 SELECTION AND MEASUREMENT OF AUDIO TEST MATERIAL

The audio material for loudness evaluation was comprised of four genres: classical, popular and jazz music, and dramatic (spoken) performance. The samples ranged in length from approximately 70 to 100 seconds, all with a loudness structure in three distinct sections:

- A beginning section of 12 to 30 seconds, recorded at a target level of -23 LUFS. This gave the listener an opportunity to establish a desired playback loudness; (the system gain structure and speaker or ear bud drive were chosen to make the beginning section softer than most listeners would prefer, which encouraged them to actively set the desired playback gain);
- A second section that dropped in loudness, lasting either approximately 25 or 50 seconds, to test whether listener accepted this loudness or desired a higher level by adjusting their playback gain;
- An ending section that equaled the loudness of the beginning section, to test listener's response to an increase in loudness that may include more playback gain added in the second section.

To reduce listener fatigue and avoid familiarity with audio selections, music and speech needed to be unique for each trial. Within each genre, audio was selected with an almost identical profile to allow data comparisons between listening environments. A total of 64 samples, 16 in each genre, were prepared. Table 2 and Table 3 list the four genres associated with each noise environments for the earbud and loudspeaker playback, respectively.

Table 2 - Test conditions with "ear bud" listening

Noise Environment		Genre Series 1				Genre Series 2			
		Classical	Speech	Pop	Jazz	Classical	Speech	Pop	Jazz
quiet office	42 dBA	C1	SP1	P1	J1	C1	SP1	P1	J1
restaurant chatter	53 dBA	C2	SP2	P2	J2	C2	SP2	P2	J2

Table 3 - Test conditions with loudspeakers

Noise Environment		Genre Series 1				Genre Series 2			
		Classical	Speech	Pop	Jazz	Classical	Speech	Pop	Jazz
quiet office	42 dBA	C1	SP1	P1	J1	C1	SP1	P1	J1
restaurant chatter	53 dBA	C2	SP2	P2	J2	C2	SP2	P2	J2
outdoor roadway	65 dBA	C3	SP3	P3	J3	C3	SP3	P3	J3
vehicle cabin	67 dBA	C4	SP4	P4	J4	C4	SP4	P4	J4

Finding samples meeting the above conditions by hand (and ear) is difficult. Consequently, a MATLAB program was developed to "crawl" through more than 3,000 music and dramatic speech recordings in NPR's Audio Library, filtering the recordings approximating the "normal-low-normal" sequence. The

[4] Each audio file was analyzed with an 8192-point Fast Fourier Transform and the values in 20 frequency bands were adjusted to the A-weighting for each band's frequency. The signal power of the 20 frequencies for each environment was summed and offset to the same value as the measured, A-weighted sound pressure levels.

resulting candidates were checked by ear and by ITU loudness charts. The chosen samples were cut from the full-length recordings at appropriate start and end points. Level of the opening and closing sections were normalized to -23 LUFS. To provide better accuracy in the listener tests, ITU loudness within the middle sections were adjusted by audio editor by up to ±2 LU to match their same-genre samples.

Figure 9 charts the loudness range test procedure using the 102 second classical sample #C2. This example includes an actual listener response using loudspeakers. In the figure, the ITU Short Term Loudness of the sample is indicated by the green line in the lower part of the graph. In the opening section of the sample, from the start to point A, the loudness is set to an average of (normalized) to -23 LUFS. The loudness of this opening section is standard for all the samples.

As in all samples, after inflection point A, the loudness drops to point B, at which the audio levels off and then continues roughly the same loudness until point C. In this sample, the mean loudness between B and C is approximately -38 LUFS – a drop of 15 LU from point A. In all samples, the loudness increases to a third section, from point D to point E, which has a mean loudness of approximately -23 LUFS. All of the inflection points were pre-determined by visual study of the loudness test samples.

For this example, the environmental noise heard by the listener was continuous, unintelligible conversation from multiple voices in a restaurant environment (but free of loud and distracting 'clanks' from plates and silverware). The average SPL of this environment was approximately 53 dBA.

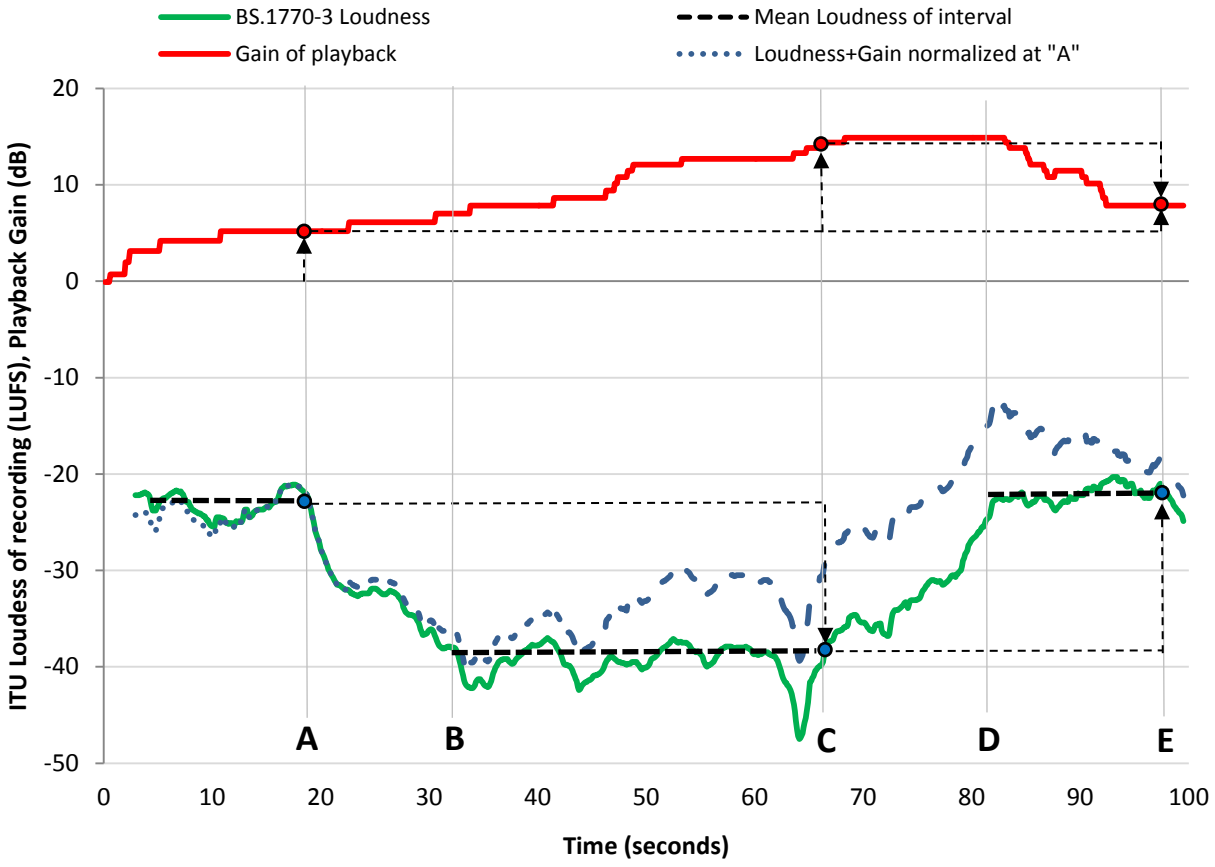


Figure 9 - Chart of an audio sample with actual listener adjustment of playback gain

The chart displays this listener's playback gain adjustments with the red line. (The gain was programmed to always start 23 dB below maximum, but is shown at 0 dB here for simplicity.) Subject #2 raised the gain by approximately 5 dB in the first 10 seconds, and then left it in place for another 10 seconds. This

subject's desired playback gain was posted at point A. After point A, the loudness began to decrease and the listener began to raise the playback gain. By the end of the lower-loudness section, at point C, this listener had increased the playback gain by approximately 10 dB, relative to the opening section at point A. Loudness for the middle section had dropped 15 LU from the opening section.

In the third section of the test, loudness of the C2 sample had returned to the nominal -23 LUFS level. This listener can be seen turning down the gain by approximately 7.5 dB, from point C to point E. The loudness level equaled the first section, and this listener had returned the playback gain to within 2.5 dB of its initial setting at point A.

8 ANALYSIS OF LISTENER TEST DATA

The gain changes represented by the dashed lines in the upper part of Figure 9, when compared to the loudness changes indicated by the dashed lines in the lower part of the figure, comprise the process of measuring listener preferences for loudness change. Participant's behavior was analyzed separately for loudspeaker listening and earbuds. Preliminary analyses showed that there were both age and gender differences, so these demographic characteristics were considered during analyses. Four analyses were conducted to determine the following behavior:

- How people adjusted their volume at the start of the listening experience to set their nominal listening level (start to point A)
- How people adjusted their volume in response to audio becoming more quiet in the middle (point A to point C)
- How people adjusted their volume in response to audio become louder at the end (point C to point E)
- The difference between their original nominal listening level and the end level after mid-sample adjustments had been made (point A to point E).

The concept of *loudness compensation* is introduced, which represents how much listeners offset the change in loudness they heard. Referring to the mean loudness points in Figure 9, the change in loudness from point A to point C is:

$$(L_C - L_A) = \big(-38 - (-23)\big) = -15\ LU.$$

Similarly, the change in listeners' playback gain at the same inflection points is:

$$(G_C - G_A) = (15 - 5) = 10\ dB.$$

Using the actual values from the figure, the loudness compensation between point A and point C combines these values as:

$$(L_C - L_A) + (G_C - G_A) = -15 + 10 = -5\ dB$$

This may be read as: "the listener's loudness compensation was 5 dB less than the decrease in loudness" (of 15 LU). If a listener fully compensated for the change in loudness, as measured by the ITU meter, the result would be 0 dB. The loudness compensation value is used in the following discussion to discuss responses to listening conditions and listener groups.

Loudspeakers

Change in playback gain at start of listening experience (start to point A)

In this opening interval, participants began by adjusting their volume to a comfortable listening level.[5] Because the opening gain was intended to play the audio softly, nearly all listeners increased the playback

[5] This is the reference stage for establishing their playback, prior to testing for loudness compensation. Hence the only measure here is their playback gain, relative to the opening gain as established by the test software.

gain, but this was mediated by their age, the type of audio (genre) and background noise. There were no differences in playback gain increase between men (+5.6 dB) and women listeners (+4.5), but older listeners increased their volume significantly more (+7.1 dB) than younger listeners (+2.4 dB). Further, listeners increased the volume less for classical music (+3.0 dB) than they did for all other genres (+5.2 dB), adjusting the volume most for speech (+5.6 dB). Finally, not surprisingly, they increased their volume more in noisier conditions (42 dBA office = +1.6 dB, 53 dBA restaurant chatter = +4.5 dB, 67 dBA cabin vehicle = +5.2 dB, and 67 dBA roadway = +7.7 dB). As would be expected, there was no difference in their adjustment based on the sample length.

Loudness compensation in response to audio becoming quieter (point A to point C)

Across all noise environments, although there was no difference between older and younger participants, women tended to compensate for the loudness change slightly more than males: -5.0 dB and -6.0 dB respectively (where 0 dB would be a full compensation for a drop in loudness). Listeners adjusted their volume less for Classical (-6.3 dB) and Speech (-6.5 dB) than they did for Jazz and Pop: -5.1 dB and -3.6 dB respectively, however male and female patterns were slightly different, as is seen in the following table:

Table 4 - Loudness compensation of male and female listeners for all noise environments (dB)

Gender	Classical	Jazz	Pop	Speech
Male	-6.8	-7.1*	-3.1	-7.1
Female	-6.0	-4.0*	-3.6	-6.3

* Significantly different at p<.001

People were influenced by background noise, such that they compensated for loudness more in noisier conditions than in quiet conditions: at 65 dBA and 67 dBA they adjusted to -3.9 dB of the actual loudness decrease, whereas at 42 dBA and 53 dBA they compensated to within -7.8 dB and -6.1 dB, respectively. This effect was affected by the genre as is seen in Table 5, which shows that Pop was compensated significantly more in noisier conditions than all other conditions:

Table 5 – Loudness Compensation with different environmental noise levels (dB)

Noise Environment		Classical	Jazz	Pop	Speech	Mean
quiet office	42 dBA	-8.7	-6.4	-4.8	-11.1	-7.8
restaurant chatter	53 dBA	-7.9	-6.0	-4.1	-6.2	-6.1
outdoor roadway	65 dBA	-4.5	-4.9	-1.8	-3.6	-3.7
vehicle cabin	67 dBA	-3.9	-3.3	-3.1	-5.2	-3.9

The length of the middle, low-level section (trough) affected people, with more compensation being made for long trough samples (-3.2 dBA) than short samples (-7.6 dBA).

Loudness compensation in response to audio becoming louder (point C to point E)

Here there was little difference in the way people adjusted volume as the loudness returned to a level equivalent to the opening section.[6] Older listeners tended to compensate slightly more vs. younger ones: 8.3 dB vs. 9.3 dB, respectively (where 8.3 dB means they under-compensated the rise in loudness by 8.3 dB) Females adjusted slightly more than males: 8.2 dB versus 9.3 dB. People adjusted classical and speech noticeably less (11.6 dB and 10.2 dB respectively) than Jazz or Pop (7.5 dBA and 5.9 dBA, respectively.) Background noise significantly interacted with genre, where listeners desired to turn the gain back down least with Classical and Speech than with Jazz and Pop (see Table 6).

[6] Note that the loudness compensation are now positive numbers because the test signals increase in loudness.

Table 6 - Loudness compensation with audio becoming louder (dB)

Noise Environment		Classical	Jazz	Pop	Speech	Mean
quiet office	42 dBA	12.0*	4.8*	6.0*	12.5*	8.8
restaurant chatter	53 dBA	12.8	10.1	6.7	10.1	9.9
outdoor roadway	65 dBA	11.5	8.5	5.6	8.4	8.5
vehicle cabin	67 dBA	10.5	7.7	6.7	10.4	8.8

* Significantly different at p<.001

The length of the trough also affected people's behavior when listening to specific genres where Classical and Speech were compensated significantly less for short troughs than long troughs; and in specific background noises, where short troughs were compensated less in quiet conditions (40 dBA) than long troughs.(See Table 7 and Table 8.)

Table 7 - Effect of length of lower-level trough on playback gain (dB)

Trough length	Classical	Jazz	Pop	Speech
Long	11.5	6.7	5.1	9.1
Short	11.7	8.9	7.3	11.5

Table 8 - Effect of length of lower-level trough versus environmental noise level

Trough length	42 dBA	53 dBA	65 dBA	68 dBA
Long	6.8*	9.6	8.2	7.9
Short	10.7*	10.3	8.6	9.5

* Significantly different at p<.001

Original gain level and ending gain level (point A to point E)

In general, people did not start and end at the same loudness, with the playback gain ending 4.8 dB higher than originally set, despite the average of the difference between last sections being only 0.3 dB lower than their respective first sections. Older people left their gain 5.6 dB higher, compared to younger people, at 3.8 dB. Classical music was most discrepant (6.7 dB), compared to Jazz (4.5 dB), Pop (3.8 dB) and Speech (4.1 dB). Background noise affected their behavior, with people choosing to leave their end-point gain closer to the point A gain in quieter environments (3.3 dB) vs. noisier environments (5.8 dB). The length of the trough did not seem to have a meaningful effect. There was a considerable range in listener opinions to restore the gain from point A to point E, a standard deviation of 7.9 dB for 1024 samples overall.

Earbuds

Change in playback gain at start of listening experience (start to point A)

Again, participants were asked to adjust their gain control to a comfortable listening level. For earbuds, there were differences between men and women listeners (+4.9 dB and +3.0 dB respectively). Further, older listeners increased their volume significantly more (+5.3 dB) than younger listeners (+1.7 dB). Listeners increased the volume less for classical music (2.4 dB) than they did for all other genres (Jazz = 3.4 dB, Pop = 4.3 dB and Speech = 4.7 dB). Finally, even with the slight increase in noise level from 42 dBA to 53 dBA, listeners increased their volume more in noisier conditions (42 dBA office = 2.4 dB, 53 dBA restaurant chatter = 5.0 dB). Again, there was no difference in their adjustment based on the length of the sample.

Loudness compensation in response to audio becoming quieter (point A to point C)

As with loudspeaker listening, there was no difference between older and younger participants, women tended to compensate for the drop in loudness slightly less than males (-4.4 dB and -5.6 dB, respectively).

However, different from loudspeaker results, listeners compensated more for Classical (-8.3 dB) than they did for Speech, Jazz and Pop (-3.0 dB, -5.1 dB and -3.0 dB respectively).

People were influenced by background noise, such that they compensated for the drop in loudness more in noisier conditions than in quiet conditions (-4.0 dB and -5.8 dB, respectively). The length of the trough affected people, with more compensation being made for long trough samples (-3.5 dB) than short samples (-6.2 dB). However, the genre had a mediating effect, such that listeners compensated for a drop in loudness significantly more in speech and jazz when the volume stayed low for a longer time period.

Table 9 - Loudness compensation relative to length of low-level trough

Trough Length	Classical	Jazz	Pop	Speech
Long	-8.7	-3.0*	-2.3	-0.1*
Short	-7.9	-7.2*	-3.8	-6.0*

* Significantly different at p<.001

Loudness compensation to audio becoming louder (point C to point E)

In contrast to loudspeaker listening, there were differences in the way people adjusted volume. Older people compensated slightly more (8.6 dB) than younger people (9.7 dB), and females compensated slightly more (8.8 dB) than males (9.5 dB). People compensated Classical music prominently less (13.8 dB) than Jazz, Speech or Pop (8.9 dB, 8.6 dB, and 5.0 dB, respectively.)

The length of the trough also affected people's behavior when listening to specific genres of music where Speech was turned up significantly more in short troughs than long troughs.

Table 10 - Loudness compensation for low-level trough duration with audio sample genres

Trough Length	Classical	Jazz	Pop	Speech
Long	14.1	8.2	4.6	7.4*
Short	13.6	9.9	5.2	9.6*

* Significantly different at p<.001

Original listening gain and ending listening gain

As with loudspeaker listening, people did not start and end at the same playback gain, but with gain ending 4.5 dB higher than originally set. Speech was most discrepant (6.5 dB), compared to Jazz (4.1 dB), Pop (2.1 dB) and Classical (5.2 dB). Although people seemed most sensitive to the original Speech level they chose with loudspeaker listening, the reverse was true for earbuds, with people seeming less sensitive to original loudness. Background noise affected this behavior, with people choosing closer end-points in quieter environments (3.6 dB) vs. noisier environments (5.3 dB). Finally, the length of the trough seemed to have a meaningful effect, particularly for speech with a long trough (see Table 11, below).

Table 11 - Effect of low-level trough duration on first- to ending-section levels

Trough Length	Classical	Jazz	Pop	Speech
Long	4.8	5.2	2.3	9.2*
Short	5.6	3.0	2.0	3.7*

* Significantly different at p<.001

9 MODEL FOR LOUDNESS MANAGEMENT BASED ON LISTENER TESTS

As discussed above, the listener tests reveal some complex relationships between loudness change and the desired "make up" in gain to please listeners. However, some generalizations emerge from the data upon which a model can be based, as shown in Figure 10 and Figure 11. In these charts, the listeners' desired changes in playback gain with loudspeakers are plotted in relation to the change in loudness, measured per ITU, from inflection points A to C.

The columns of points indicate listener responses in common with various audio samples. The four noise environments are divided into two charts for clarity: office and roadway representing the lowest and highest noise environments, respectively, while the restaurant and in-vehicle noise are in between.

Figure 12 provides the listener's change in gain versus change in loudness for earbuds. As explained earlier, the two highest noise environments were excluded as a precaution against poor data.

For each figure, there are a number of points on the 0 dB line, indicating that some listeners decided that their initial gain setting was fine across the low-loudness trough. While it could be argued that no loudness management is needed for these listeners, their scores were considered in the following analysis. (There are also a small number of gain scores in the negative range, meaning that the listener preferred to turn down their playback gain, in opposition to the reduced loudness. Their data were too infrequent to affect the overall results.)

For each noise environment, a linear regression was performed on the data: the slope and intercept are shown adjacent to each trend line. The upward slope to the right indicates that as the loudness change increases, the listener means for playback gain also increase.

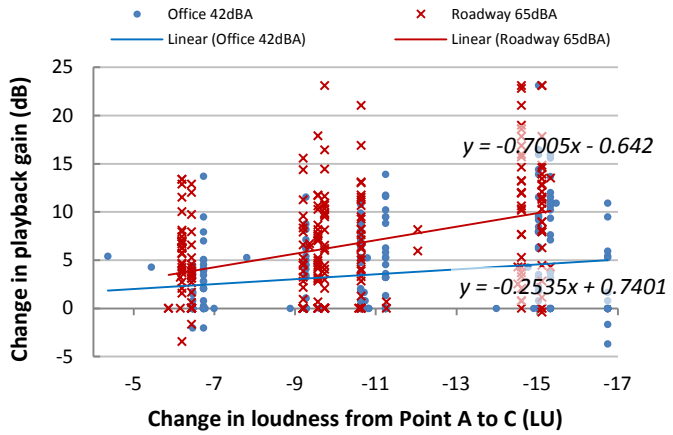


Figure 10 - Change in loudness for all listeners and all genres with office and roadway noise environments using loudspeakers

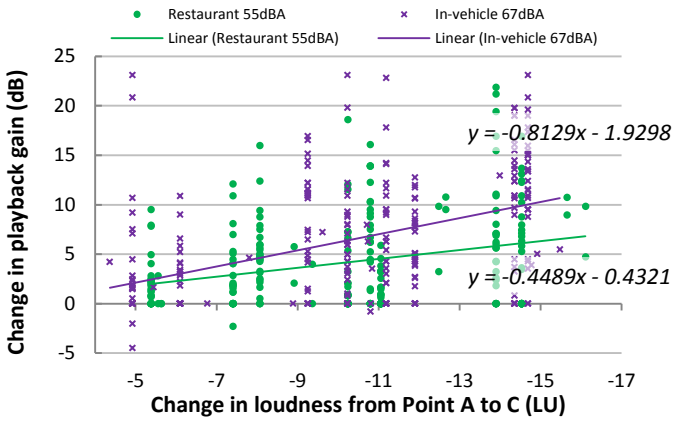


Figure 11 - Change in loudness for all listeners and all genres with restaurant and in-vehicle noise environments using loudspeakers

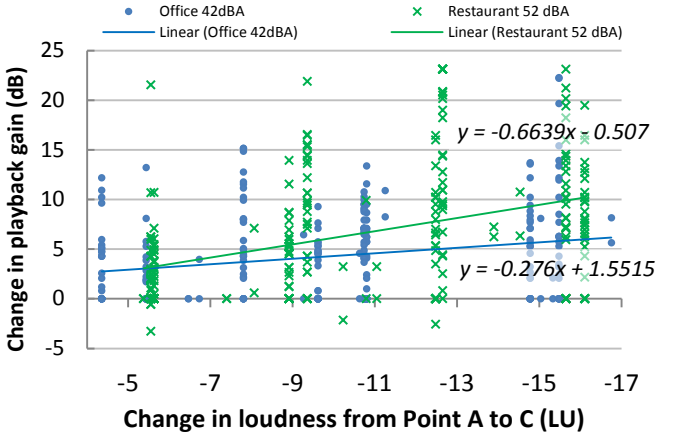


Figure 12 - Change in loudness for all listener and all genres with office and restaurant noise environments using earbuds

Table 12 - Comparison of compensation slopes for playback gain in various noise environments (all genres)

Listening Mode	Noise Environment	SPL (dBA)	Slope	Intercept	Slope with zero intercept
Loudspeaker	Quiet office	42	-0.2535	+0.740	-0.3124
	Restaurant	53	-0.4489	-0.432	-0.4100
	Roadway	65	-0.7005	-0.642	-0.6427
	In-vehicle	67	-0.8129	-1.929	-0.6432
Earbud	Quiet office	42	-0.2760	+1.551	-0.4123
	Restaurant	53	-0.6639	-0.507	-0.6219

As shown in Figure 10 through Figure 12, as the noise environment level increases, the slopes increase. Thus, loudness management, or the desired compensation of loudness variation, is based on two elements:

(1) Measuring the change in loudness by an appropriate system such as specified by the ITU, and

(2) Active sensing the level of ambient noise in the listening environment by microphone.

To convert the results into a form that is practical for loudness management, the slopes must be considered with the noise environments. The intercepts complicate the reduction, but fortunately are relatively small in magnitude (+1.5 to -0.6 dB). The regression can be set to intercept with 0, a noise-free condition for which no loudness management is assumed necessary. Table 12 lists the slopes with zero intercept, which are plotted as a group in Figure 13.

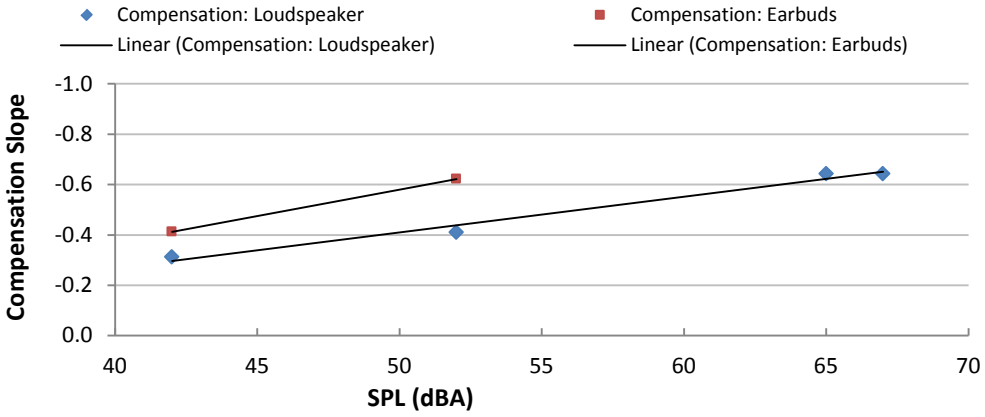


Figure 13 - Chart illustrating the compensation slopes as a function of environmental noise level

To apply the result, consider an ambient noise level of 50 dBA. The compensation slope for loudspeakers is approximately -0.40, meaning that if the audio program loudness changed by -10 LU, the compensation would be: -10 x -0.40 = +4 dB, an increase in gain of 4 dB under these circumstances. At an ambient noise level of 65 dBA, the compensation for a change of -10 LU would be: -10 x -0.62 = +6.2 dB. Being a simple algorithm, this automatic process could be added to the DSP volume control in a range of consumer audio playback equipment, including portable devices and automobiles.

10 CONCLUSIONS

While the data represent the central tendency of listeners, there are listeners who desire more or less control than the mean gain. It is recognized that no "one size fits all" solution exists, but there is a consistent and usable trend. Depending on the listener's tastes, an "activity" control could be included to change the slope, so that more or less compensation could be applied to the changes in audio loudness.

As the data indicate that speech needs more compensation that music, the compensation could include a switch for either program type.

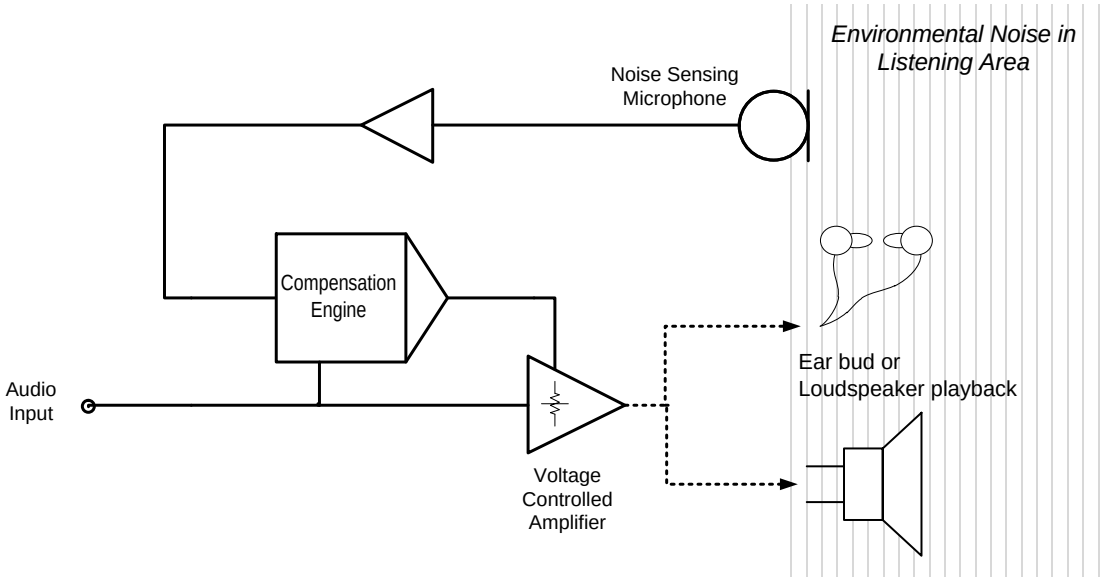


Figure 14 - Simplified diagram showing how the compensation algorithm could be implemented in a playback device.

Figure 14 illustrates the basic components of the loudness compensation system within the playback device, such as a smartphone or home stereo. The noise sensing microphone picks up the noise field near the listener and presents the noise signal to the Compensation Engine, which determines the sound pressure level and applies a slope factor to produce a gain control signal for the voltage controlled amplifier. For loudspeaker playback, a linkage from the input audio is shown, which could freeze the environmental noise measurement once program audio begins, which could be misinterpreted as environmental noise. The microphone is optional, however, if a control were available so that the listener could set the average conditions, such as "quiet office" or "noisy home" to effect the appropriate loudness compensation.

The results indicate that the adjustment is not a dB-for-dB compensation to the change in loudness. Listeners do not require a high "compression ratio" (to put it in terms of audio processor design) for control of most program audio. Apparently, they do not wish a constant "signal-to-noise ratio" in listening to program material with various levels of ambient noise. (Another way to look at it may be that they have limits on how loud they wish the program material to be, and they consciously or subconsciously anticipate that program material could become louder at a later time.)

It should be noted that most program material delivered to consumers has undergone dynamic processing, possibly at more than one stage from studio production to mass-distribution. The material selected for listener testing was chosen to have loudness variation, however, much of the audio content sold online, mass-produced on CD or delivered by radio and television has minimal loudness range. In these cases, a loudness management system of the type described herein would likely set playback gain relative to the ambient noise and "sit out" further changes until the program content varied.

This study looked at decreases in loudness, not unexpected increases due to mismatched levels in audio content. As discussed at the outset, this is another issue that deserves attention. The solution for the management of inconsistent levels across different content could easily be integrated into the described system.

This study suggests further study of loudness measures. For example, the ITU loudness meter provides a "Momentary" measurement with an integration time of 0.3 s, in addition to the Short Term measurement

of 3 seconds used in this study. The correlation of faster Momentary values to listener data may reveal more about the speed at which listeners respond to changes in loudness – something beyond the scope of this study. The ITU standard's Loudness Range (LRA) is a relatively long-term measure, which could provide interesting results in comparison to listener responses. The CBS Loudness Meter is a different measure of loudness that could be compared in listener response to the ITU meter. [7] This study is an initial investigation into the loudness preferences of average listeners. Further research would add to the approach and optimize it for specific playback systems.

11 ACKNOWLEDGEMENTS

The authors wish to thank NPR Labs intern Jeremy Adams, who recruited, scheduled and supervised most the listeners for the testing, and technical researcher Alice Goldfarb for processing the data tables. We greatly appreciate the help of David Berg at Oldfield Laboratories, Inc. in Minneapolis for providing test data on the Apple Ear Pods used in the listening tests.

12 REFERENCES

[1] D. Giannoulis, M. Massberg and J. Reiss, "Digital Dynamic Range Compressor Design—A Tutorial and Analysis," J. Audio Eng. Soc., Vol. 60, No. 6, 2012.

[2] The ITU Radiocommunication Assembly (ITU-R), "Recommendation ITU-R BS.1770-3, Algorithms to measure audio programme loudness and true-peak audio level," August 2012. [Online]. Available: http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-3-201208-I!!PDF-E.pdf. [Accessed June 2014].

[3] E. Sheffield and J. Kean, "Streaming Codec Study Report for NPR Digital Media," January 2013. [Online]. Available: http://nprlabs.org/sites/nprlabs/files/documents/codec/20130920 Streaming Codec Study Report.pdf. [Accessed June 2014].

[4] International Standards Organization, (ISO), "ISO 226:2003, Acoustics - Normal equal-loudness-level contours," 05 2014. [Online]. Available: http://www.iso.org/iso/catalogue_detail?csnumber=34222. [Accessed 22 10 2014].

[5] D. Berg, "iPhone 5 ear bud measurements," for NPR Labs, courtesy of Oldfield Labs, Minneapolis MN, 2014.

[6] InnerFidelity, "Headphone Measurements: Apple iPod Ear Buds sample A," [Online]. Available: http://www.innerfidelity.com/images/AppleiPodEarBudssampleA.pdf. [Accessed 11 2014].

[7] B. L. Jones and E. L. Torrick, "A New Loudness Indicator for Use in Broadcasting," *SMPTE Journal*, pp. 772-777, 1981.