Audio Engineering Society

Convention Paper 8821

Presented at the 134th Convention
4th-7th May, 2013 - Rome, Italy



# A Framework for Adaptive Real-Time Loudness Control

Andrea Alemanno[1], Alessandro Travaglini[2], Simone Scardapane[1], Danilo Comminiello[1] and Aurelio Uncini[1]

[1] Department of Information, Electronic and Telecommunication (DIET) - University of Rome "La Sapienza", I-00184 Rome – Italy
{andrea.alemanno, simone.scardapane, danilo.comminiello}@uniroma1.it, aurel@ieee.org

[2] Fox International Channels Italy, Rome, I-00138, Italy
alessandro.travaglini@fox.com

## ABSTRACT

Over the last few years, loudness control represents one of the most frequently investigated topics in audio signal processing. In this paper we describe a framework designed to provide adaptive real-time loudness measurement and processing of audio files and streamed content being reproduced by mobile players hosted in laptops, tablets and mobile phones. The proposed method aims to improve the users listening experience by normalizing the loudness level of the content in real-time, while preserving the original creative intent of the original soundtrack. The loudness measurement and adaptation is based on a customization of the High Efficiency Loudness Model algorithm described in the AES Convention Paper #8612. Technical and subjective tests were performed in order to evaluate the performance of the proposed method. In addition, the way the subjective test was arranged offered the opportunity to gather information on the preferred Target Level of streamed and media files reproduced on portable devices.

## 1. INTRODUCTION

One of the most frequently reported complaints to television operators is the annoyance generated by programmes being reproduced at inconsistent loudness levels. In the past few years, the international broadcast community has deeply investigated this issue and since then many solutions have populated the market, including hardware and software, real-time or file-based products. The webcast industry, which in the last few years has enormously increased its content delivery service, seems to suffer from the same problems broadcasting reported until a short time ago.

This paper analyzes this scenario and describes the development of a framework designed to cope with the need of measuring and controlling loudness levels of file-based and streaming content in real-time. In order to verify the performance of the application, several stimulus of different genres and dynamics were used and the resulting waveforms were analyzed.

Moreover, a subjective MOS test was arranged. The topics presented in the paper are arranged as follows: challenges of real-time loudness control in Section 2, overall architecture of the framework in Section 3, algorithm used for loudness measurement and the loudness controller design in Sections 4 and 5 respectively, subjective and objective tests results in Section 6, and finally conclusions and guidelines for further developments in Section 7.

## 2. REAL TIME LOUDNESS MEASUREMENT

In broadcasting, processing loudness levels of live programmes presents a particular challenge due to the tight relations between the original creative intent, the required loudness characteristics, and the listening comfort ultimately provided to the viewer. These issues also occur when processing loudness levels of streamed content, where several implications occur. In particular the source material originates from a much less controlled workflow, the range of possible original levels and dynamics is very large, and the computational resources are much weaker than those available in professional environments. In fact, a software for controlling loudness levels in real-time should simultaneously fulfill the following requirements: present the audio programmes at a consistent level, preserve the original creative intent as initially produced, measure and process the loudness levels as close to real-time as possible, and fulfill the user expectations according to the restricted listening comfort provided by PC and mobile devices.

The model described in this paper has been designed bearing in mind these goals. Whilst some existing tools are designed for being hosted on TV sets and audio amplifiers (i.e. Dolby Volume), at the time of writing the authors are not aware of any other real-time loudness processors that can achieve these goals simultaneously. Streamed content is transmitted with very inconsistent levels, unless normalized at station level or by specific tools present in the reproduction device (PC, tablet or mobile phone). File-based content stored in the user's medium, typically the media player needs to analyze and normalize all of the files before the user can listen to it. This is the main obstacle because in order to apply a gain control to a huge playlist consisting of thousands of tracks the user has to wait an undefined time, necessary to let the algorithm analyze the entire tracks list. The model presented in this paper addresses this problem and allows the user to play any track in real-time without having to process it entirely in advance. For the intent of presenting the designed solution in this paper the framework was developed only as a media player. Further work is in progress to extend its features for implementation in webcasting where the exact same processing would be applied.

## 3. FRAMEWORK ARCHITECTURE

Bearing in mind the objectives described in Section 2, the overall architecture of the framework was designed in order to achieve efficiency, simplicity and robustness. Figure 1 shows an overview of the implementation, which is split into three main components, labeled *media player*, *real-time loudness processor*, and *adaptive gain controller.*

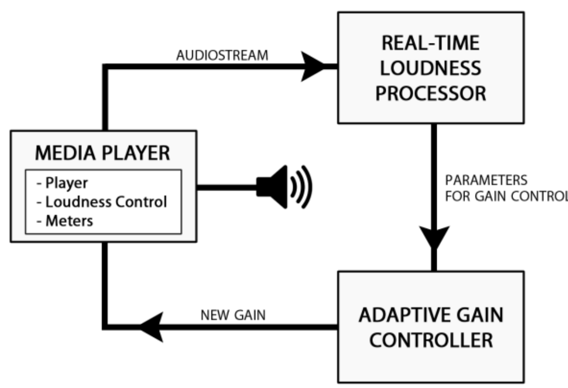


Figure 1 Framework Architecture

- MEDIA PLAYER: The framework includes a media player used to select the track and to operate on the main Transport, Volume and Gain controls. The user is free to decide whether he/she wants to listen to the track at its original level and dynamics or if he/she wants to enable the Loudness Control. The media player controls are organized in three distinct panels, respectively the player panel, the loudness control panel and the meters panel, as shown in Figure 2.

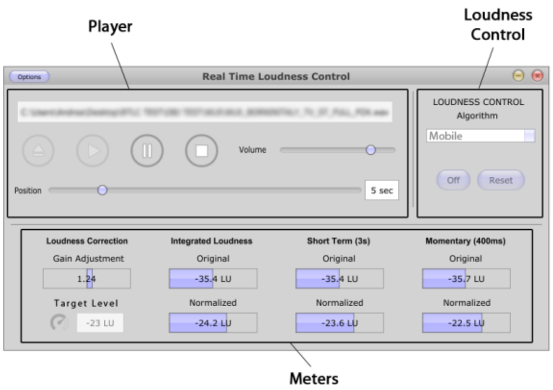


Figure 2 Controller GUI

- The *player panel* includes 4 Transport buttons and 2 faders, one for the Volume Control and one to control the relative listening level of the track being played.

- The *loudness control panel* allows the user to enable, reset and disable the real-time loudness controller.

- The *meters panel* presents a series of indicators that show several parameters described in EBU R128 [1] and EBU TECH 3341 [6] and include the Integrated Loudness Level, the Short Term Loudness Level and the Momentary Loudness Level of both the original track and the real-time processed version. These six meters become active only when the Loudness Control feature is turned on. On the left-hand side of the meters area another indicator shows the real-time overall gain adjustment and a rotary slider shows the Target Level selected by the user before playing a track.

• REAL-TIME LOUDNESS PROCESSOR: The implemented real-time loudness processor is the main component of the framework. When the user enables the loudness control in the media player, the processor initiates the real-time measurement and the corresponding gain control is applied. The measurement algorithm implemented is a variation of the High Efficiency Loudness Model described in AES Conference Paper #8612 [2] (see Section 4). While the main structure has been left as originally designed, in order to better emulate the audio characteristics of the loudspeakers typically implemented in mobile devices, the weighting filters have been remodeled. Furthermore, the Adaptive Loudness Processor has been designed to be as less computationally intensive as possible, with a strong emphasis on avoiding unnecessary latency.

• ADAPTIVE GAIN CONTROLLER: The adaptive control is divided in two parts: 1) A pre-alignment measurement based on the Integrated Loudness value calculated right before the track is played back; 2) an adaptive loudness measurement constantly updated during playback. The adaptive gain controller is discussed in depth in Section 5.

## 4. MEASUREMENT ALGORITHM

The framework performs the measurement using a small variation of the algorithm presented in Budapest at the 132th AES Convention in the CP #8612 “HELM: High Efficiency Loudness Model”. A brief review of its design is given below.

The original algorithm uses a pre-filtering model, based on several psychoacoustics studies (Blauert [3] e Moore [4]). In this adaptive application it is modified in order to improve its performance for real-time analysis. As mentioned above the weighting filters have been modified in order to better emulate the typical mobile device loudspeakers frequency response, including laptop computers, tablets and cell phones.

After the frequency weighting, the algorithm performs the following operations in sequence: (i) the power computation; (ii) the relative threshold computation based on the absolute threshold; (iii) and finally the Integrated Loudness computation (or Program Loudness if the entire signal has been analyzed). The measurement applies a 400ms rectangular sliding window with 75% overlap. Differently to its original version, the Adaptive Processor implements Relative Gating [2] as oppose to Recursive Gating because of the real-time needs that this application requires.

## 5. ADAPTIVE GAIN CONTROLLER

Adaptive Gain Controller is the main core of the framework. Working with a real-time adaptive gain control based on the Loudness measurement presents a wide range of issues that have to be addressed in order to achieve an efficient and comfortable listening experience. The control is split into two main operations: pre-buffering processing, and real-time processing.

In the former a look-ahead estimation of the first few seconds of Integrated Loudness Level is computed. In the final implementation, this operation takes typically about 100ms.

Then, according to this early measurement, a pre-gain is performed in order to provide an early alignment of the track with the Target Level set by the user. This operation is required to achieve a very reactive yet efficient loudness normalization, and regardless to the genre, duration, content and spectral balance of the soundtrack. While providing a very responsive overall normalization this operation also allows the gain controller to set its operation around the Target Level, thus optimizing the performance of the dynamic processing. In the consequent real-time processing, while the track is playing, the controller keeps adjusting the gain according to the real-time computation of the Short Term Loudness performed in the background by the framework.

The gain control is carried out by a series of operations that seek to keep a comfortable balance between the original sound dynamics of the track and the actual Target Loudness. The processing architecture of the framework guarantees that the technical and the aesthetical requirements are always fulfilled. Currently, the algorithm is always set to reach the same balance between the original intent of the track and the need to follow the Target Level.

A future development, already in progress, will let the user decide how much processing shall be applied in regard to the original dynamics, and to the Target Loudness Level selected. This feature will allow the user to accommodate his/her preferred listening taste.

## 6. TEST PHASE

In order to prove the effective robustness of the model we performed both objective and subjective tests.

### 6.1. Objective Test Set-up

In the objective test we analyzed two loudness parameters: the Integrated Loudness Level, and the Short Term Loudness Level. The test was executed on several types of content with different audio characteristics: Wide Loudness Range, Narrow Loudness Range, Speech-based, Music-based, and Heavy Low Frequency tracks.

For all these categories we used audio tracks gathered from the official EBU-PLOUD database. In order to assess the performance of the framework we compared the original Loudness values of these tracks with the levels resulting after processing. Results are shown in Section 7.1.

### 6.2. Subjective Test Set-up

The subjective test was performed using a Samsung laptop computer model Series 3.

For acquiring the subjective assessments we used a small variation of the Mean Opinion Score (MOS) procedure [5]. This is an effective procedure that gives the possibility of numerically estimating a subjective measure.

The subject, who was not allowed to see the screen nor to receive visual information from the application, was required to listen to the tracks and to give a rating, using the scale shown in Table 1. The final MOS value is the arithmetical mean of all the individual scores gathered, ranging from 1 (worst) to 5 (best).

| MOS | QUALITY | IMPAIRMENT |
|---|---|---|
| 5 | Excellent | Imperceptible |
| 4 | Good | Slightly perceptible but not annoying |
| 3 | Fair | Slightly annoying |
| 2 | Poor | Annoying |
| 1 | Bad | Very annoying |

Table 1 MOS Ranking

In our MOS test we prepared a playlist of 12 tracks, taken from the same database as for the objective test, this time including the following categories: WLR (wide loudness range), NLR (narrow loudness range), speech, music, Generic (music + speech), HLF (tracks with heavy low frequencies content).

Two tracks from each category were arranged in the playlist in order to test the implementation on multiple content types. Before proceeding with the test we modified the audio level of these tracks in order to recreate a very inconsistent playlist. Every track has been modified randomly with a gain correction in the range of ±10dB. The test was performed in the following order:

- A short introduction to the test was given to the subject
- The subject filled in a quick form with some general information
- The subject played a reference track and adjusted the Target Level so as to simulate his/her typical preferred listening level
- The subject started listening to the shuffled playlist, with the gain controller set OFF
- The subject answered the first round of MOS questions
- The subject started listening to the shuffled playlist, with the gain controller set ON
- The subject answered the second round of MOS questions

It is important to note that the subject performed the test blindfolded, without knowing if he/she was listening to the normalized or the original playlist, thus ruling out unwanted influences.

The total running time of the test was on average approximately 12 minutes, and never exceeded 15 minutes. The test was performed individually.

## 7. PERFORMANCE EVALUATION

### 7.1. Objective Test Results

For the Integrated Loudness Level we evaluated how close to the Target Level the processed files were normalized.

Regardless to the original track level, results indicate that in most cases the framework can efficiently normalize the whole programme within a small tolerance of ±1,0 LU. On a few occasions the tracks have been normalized within a slightly larger range (±2,0 LU). No normalization exceeded this range.

For the control of the Short Loudness Levels the objective tests highlight that the loudness processor is capable of maintaining the upper loudness modulation of the tracks within a few LU (typically 5LU), thus providing a comfortable presentation of the audio programmes, by eliminating any annoyance for the listeners.

This result is guaranteed for all tracks, regardless their genre or audio characteristics. The curves shown in Fig. 3-7 compare the original and the resulting Short-Term modulations for one representative track of each category.

As can be seen, whilst the loudness ranges are controlled, the original creative intent is preserved as the lines follow the same trends, while the processed programmes are presented within more comfortable margins.

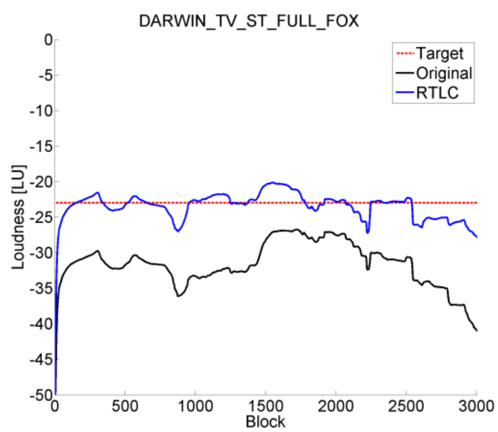


Figure 3 - WLR Content – STL

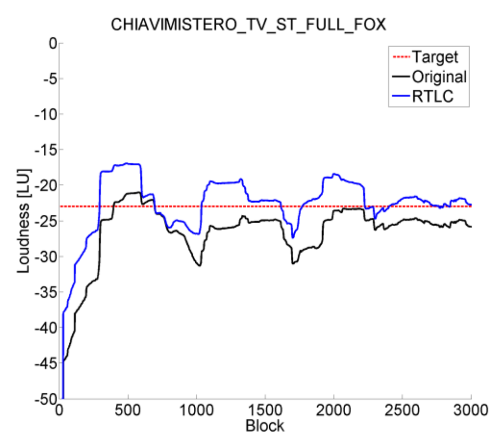


Figure 4 - NLR Content – STL

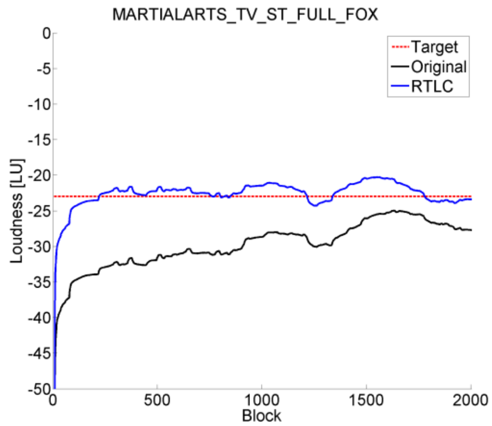


Figure 5 - Music Content – STL

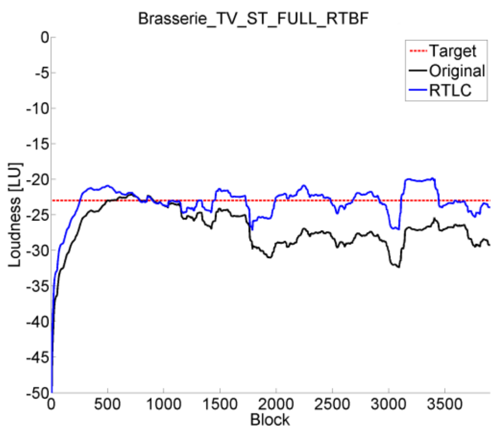


Figure 6 – Speech Content – STL

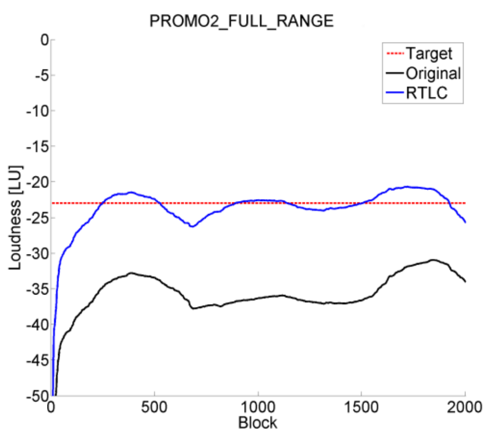


Figure 7 - Weighting Issue Content – STL

### 7.2. Subjective Test Results

The subjective tests provided information on how effective and pleasant the loudness control is. From the results shown in Tables 2 and 3 it emerged that the performance of the framework was highly appreciated by the subjects, with a final MOS average for the processed files of 4.3 (corresponding to quality assessment between "good" and "excellent") versus a MOS score for the original tracks of 1.8.

Tables 2 and 3 report respectively the entries provided by each subject and the total result for the MOS test performed.

| Subj. | MOS | | TARGET LEVEL PREF. |
|---|---|---|---|
| | ORIGINAL | NORMALIZED | |
| **1** | 1 | 5 | -20 |
| **2** | 2 | 4 | -19 |
| **3** | 1 | 4 | -21 |
| **4** | 2 | 3 | -19 |
| **5** | 2 | 4 | -20 |
| **6** | 1 | 4 | -18 |
| **7** | 3 | 5 | -20 |
| **8** | 3 | 4 | -17 |
| **9** | 1 | 5 | -20 |
| **10** | 2 | 5 | -18 |
| **11** | 3 | 5 | -19 |
| **12** | 1 | 5 | -20 |
| **13** | 2 | 4 | -20 |
| **14** | 1 | 5 | -18 |
| **15** | 2 | 4 | -18 |
| **16** | 2 | 4 | -18 |
| **17** | 2 | 4 | -20 |
| **18** | 3 | 5 | -18 |
| **19** | 1 | 4 | -19 |
| **20** | 2 | 4 | -20 |
| **21** | 2 | 4 | -18 |
| **22** | 2 | 4 | -21 |

Table 2 - MOS Results and Target Level Preference for each subject

| MOS | | TARGET LEVEL |
|---|---|---|
| ORIGINAL | NORMALIZED | |
| 1.8 | 4.3 | -19.1 |

Table 3 - Total MOS Test and Target Level Result

In addition, the subjective test gave us the opportunity to gather information on the preferred Target Loudness Level of programmes being played through a computer. The authors believe this information can be useful in setting guidelines and technical recommendations for loudness in webcasting, streaming and other portable medium. The resulting preferred Target Loudness Level was -19LUFS.

## 8. CONCLUSIONS

In the last few years loudness normalization has represented one of the most investigated issues in audio signal processing. Whilst broadcasting has already implemented several technical solutions aimed at addressing the problem, in other medium such as webcast streaming and file-based mobile players, at the time of writing this paper, not enough evidence appears to be gathered in order to provide similar solutions.

The aim of the research described in this paper was to design a framework capable of providing adaptive real-time loudness measurement and processing of either media files and streamed content being reproduced by mobile players running on laptops, tablets and mobile phones. The proposed application aims to provide a comfortable listening experience to users by normalizing the loudness level of the content, while preserving the original creative intent of the original soundtrack.

Objective and subjective tests were performed in order to verify the efficiency and the robustness of the designed framework. Both tests highlighted very good performance that encourage us in developing the application further. In addition, the subjective test gave the opportunity to gather information on the preferred Target Level which resulted in the value of -19LUFS.

However, considering that this alignment was obtained with the PC master level set at its maximum value (100) we would recommend to scale the Target Level a few LU higher, in a value ranging between -16 and -14 LUFS in order to leave some headroom for further user-end amplification.

## 9. ACKNOWLEDGEMENTS

The authors would like to thank Debbie Samuels for her assistance in editing this paper.

## 10. REFERENCES

[1] EBU Technical Recommendation R 128 “Loudness normalisation and permitted maximum level of audio signals” (2010)

[2] AES Convention Paper 8612 – Alessandro Travaglini, Andrea Alemanno, Aurelio Uncini - “HELM: High Efficiency Loudness Model for Broadcast Content” (2012)

[3] JENS BLAUERT – “Spatial Hearing: The Psychophysics of Human Sound Localization” (1997)

[4] BRIAN C. J. MOORE – “An introduction to the Psychology of Hearing” - Fifth Edition (2004)

[5] ITU-T Recommendation P.800 – “Methods for subjective determination of transmission quality” (1996)

[6] EBU TECH 3341 - Loudness Metering: ‘EBU Mode’ metering to supplement loudness normalisation in accordance with EBU R 128 (2011)