Audio Engineering Society

Convention Paper 8836

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

Automated Tonal Balance Enhancement for Audio Mastering Applications

Stylianos - Ioannis Mimilakis[1], Konstantinos Drossos[2], Andreas Floros[2], and Dionysios T. G. Katerelos[1]

[1] *Dept. of Sound & Musical Instruments Technology, Technological Educational Institute of Ionian Islands, Lixouri, Greece*

[2] *Digital Audio Processing & Applications Group, Dept. of Audiovisual Arts, Ionian University, Corfu, Greece*

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

ABSTRACT

Modern audio mastering procedures are involved with the selective enhancement or attenuation of specific frequency bands. The main reason is the tonal enhancement of the original / unmastered audio material. The aforementioned process is mostly based on the musical information and the mode of the audio material. This information can be retrieved from a listening procedure of the original stimuli, or the corespondent musical key notes. The current work presents an adaptive and automated equalization system that performs the aforementioned mastering procedure, based on a novel method of fundamental frequency tracking. In addition to this, the overall system is being evaluated with objective PEAQ analysis and subjective listening tests in real mastering audio conditions.

1. INTRODUCTION

Commercial audio production consists of several stages and often, if not always, the final one is called mastering. It involves the process where the overall audio enhancement and modification takes place just before the stage of replication and distribution [1, 2] and it aims to link the perceived characteristics of the professional audio production industry with the hi-fidelity/home-entertainment ones [1]. To this cause, technical and aesthetic approaches for audio equalization and compression have been widely employed, in order to maximize the perceived loudness and to provide the required audio content overall enhancement [1].

Focusing on the equalization process, typical mastering techniques include a tonal balancing step in order to reduce unwanted effects, like the spectral masking produced by the mixture of sources in a recording [2, 3]. More specifically, this operation implies an enhancement of specific frequency bands, according to the audio material's musical key. This process results into a spectral modification of the aforementioned frequency bands, where the corresponding note of the musical key lies, together with its even and odd harmonics [2]. The above musical key identification is performed subjectively by the audio engineer through continuous listening, or directly from the corresponding musical score (if available).

Nevertheless, in the majority of musical pieces the key tends to alter frequently. This fact imposes a continuous human effort and time-consuming operations for the realization of the required spectral processing. Towards the aim of a possible time-cost reduction for a mastering engineer to fulfill the aforementioned tasks, emerging automated and adaptive techniques seem to be attractive alternatives. To this end, pitch trackers are likely to be a promising choice, since they can provide a robust fundamental frequency estimation.

Recently, a pitch tracker was introduced that meets the aforementioned criteria of automation, adaptation and robustness [4]. Moreover, it's time-domain implementation induces significantly decreased computational cost, rendering it suitable for real-time operation and providing a feasible and effective solution for the task of fundamental frequency estimation. In this work, we propose an adaptive and automated tonal balance enhancing system for audio mastering applications using the above pitch tracker and a set of second-order peaking type filters. The central frequency of the latter is adjusted by the most prominent frequencies derived from a frequency histogram obtained by the tracker's output. The overall system efficiency is assessed using a series of subjective (listening) tests and objective audio analysis.

The rest of the paper is organized as follows: in Section 2 a detailed overview of the proposed system architecture is provided. Section 3 includes a presentation of the experimental setup followed, focusing on both the objective and subjective tests. Section 4 demonstrates the perceptual efficiency of the proposed system, while Section 5 concludes the work and determines some potential improvements that can be considered in the future.

2. SYSTEM OVERVIEW

The proposed system's architecture can be summarized as illustrated in Figure 1. In particular, it consists of two subsystems: a) the Pitch Tracker, and b) the Equalization Subsystem. The input audio signal is duplicated and each of the two copies serves as input for both subsystems.

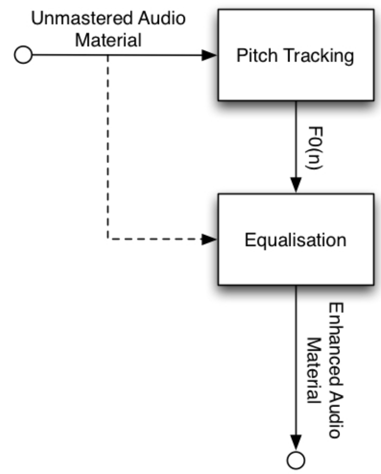

Fig. 1: *Proposed system's architecture.*

2.1. The Pitch Tracking Subsystem

As it is shown in Figure 2, the first copy of the unmastered audio material is being used to provide pitch estimates (in terms of the fundamental frequency $F0$), using frequency demodulation based on a third-order phase-lock loop (PLL) [4]. The demodulator employed here is capable of providing sample-by-sample pitch estimates, using a loop filter that operates in non-linear mode. In order to compute the exact pitch values, a pre-processing stage must be applied in order to enhance the tracking ability of the PLL system. For further analysis on the PLL system used, the reader is encouraged to refer to [4].

The aforementioned pre-processing stage consists of monophonic summation, spectral band-limiting in the region of 20Hz and 1.5kHz and constant gain envelope weighting as expressed by Equation 1 [4]:

$$x_{in}(n) = [in_{(M)}(n) \times h_{LP}(n) \times h_{HP}(n)] \cdot g_{env}(n) \quad (1)$$

where $x_{in}(n)$ is the input of the pitch estimator,

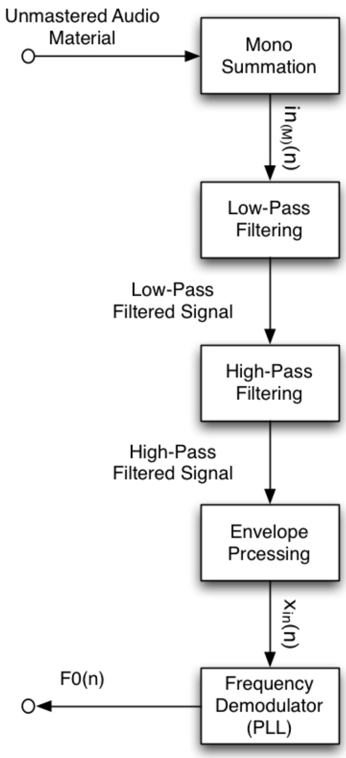

Fig. 2: *Pitch Tracking Subsystem.*

$in_{(M)}(n)$ is the monophonic summation of the original audio material, $h_{LP}(n)$ and $h_{HP}(n)$ are the 6th order low-pass and 2nd order high-pass filters provided by [5] and $g_{env}(n)$ is the envelope gain factor imposed by the inversion of the time domain band-limited signal $g_{env}(n) = 1/x_{env}(n)$, with x_{env} representing the envelope of the input signal.

The main reason for applying monophonic summation is that the information of both channels should be used for the pitch computation, since the same equalization must applied to each channel in order to preserve the original mixing balance [2]. To do so, the method of averaging the two channels was used, since it can yield good results in a simple case of stereophonic audio material [6].

After the above pre-processing stage, the input signal x_{in} is served as input to the actual pitch estimation module. At this stage, x_{in} is multiplied by a feedback oscillator output and the loop gain K_d of the present feedback system which controls the frequency range of detection [4].

The feedback oscillator is managed by the output of a second-order loop filter and can be implemented as a recursive complex valued filter with quadrature output signals, which cosine part is used for providing estimates for $F0(n)$.

The aforementioned loop filter is a combination of a second-order low-pass filter and a direct path, which forms a low-frequency shelving filter with a quality factor Q_{pll}, providing parametric and robust pitch tracking for audio signals and leading to a third-order PLL [4].

2.2. Equalisation Subsystem

The second copy of the original audio material and the $F0(n)$ estimates extracted by the pitch tracker serve as inputs to the second subsystem implementing the equalization process. A preview of this subsystem architecture is appeared in Figure 3.

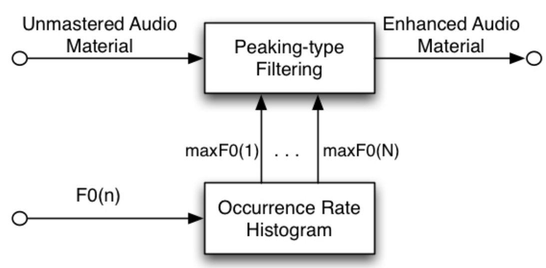

Fig. 3: *Equalisation Subsystem.*

In detail, the derived $F0(n)$ estimates are used to create a histogram of occurrence for each frequency within the 20Hz - $1.5\,kHz$ frequency band. The frequencies at which histogram peaks are observed ($max_{F0}(N)$) are selected and considered as centre frequencies (f_c) for a set of second order peaking-type filters. Each channel of the unmastered audio material is then fed to the subsystem and filtered respectively, following the analysis presented in [5] and using the equations:

$$x_h(n) = x(n) - d(1-c)x_h(n-1) + cx_h(n-2) \quad (2)$$

$$y_1(n) = -cx_h(n) + d(1-c)x_h(n-1) + x_h(n-2) \quad (3)$$

$$y(n) = \frac{H_0}{2}[x(n) - y_1(n)] + x(n) \quad (4)$$

In the above formulas d denotes the controller of the central frequency, i.e. $d = -cos(2\pi \frac{f_c}{f_S})$. Moreover, $x(n)$ is the input signal, $V0 = 10^{\frac{G}{20}}$ is the gain factor, the parameter $H0$ equals to $V0 - 1$ and f_b is the normalised bandwidth which is controlled by the boost (c_{boost}) and cut (c_{cut}) coefficients derived from Equations 5 and 6 respectively as:

$$c_{boost} = \frac{tan(\pi f_b/fs) - 1}{tan(\pi f_b/fs) + 1} \quad (5)$$

$$c_{cut} = \frac{tan(\pi f_b/fs) - V_0}{tan(\pi f_b/fs) + V_0} \quad (6)$$

3. EXPERIMENTAL SETUP

A realistic scenario for the proposed system's performance assessment should include subjective evaluation from human specialists that are trained to perform audio mastering tasks. This is due to the fact that these kind of procedures are often, if not always, empirical in nature and are based on the aesthetics of the mastering engineer in-charge. Moreover, an additional stage of objective evaluation has been performed in order to evaluate the proposed system's performance in terms of the perceived quality, clarity and fidelity. More details are provided in the following two sub-sections.

For all the test sequences employed (subjective and objective), we considered four (4) stereo master music tracks, obtained from the samples library available at [7]. The title, artist name and the genre of these tracks are shown in Table 1. A musical genre range from electronic to rock music was selected in order to evaluate the system's performance in various and typical musical styles. It should be also noted that from the great variety of samples provided by [7], we selected those that their genre categorization is not a combination of two or more sub-genres (i.e. Pop/Jazz etc). Finally, for each of the above tracks, an enhanced audio version was created, using the tonal balancing system presented in this work.

For the tonal balancing algorithm we used the parameter values summarized in Table 2. The selection of the peaking-type filters' gain factor(G) and bandwidth(f_b) is based on optimal values commonly used in similar procedures according to [1]. As far as the parameters controlling the pitch tracking functionality are concerned, the loop-filter's gain K_d value of 0.12 provides a feasible solution for audio signals that have already been filtered by a low-pass filter with a cut-off frequency of 1.5kHz [4]. In addition, a non-overshooting value of 1/3 of the quality factor of the low-shelving filter Q_{PLL}, defined within the frequency demodulator, can provide more detailed information about the exact state of the fundamental frequency [4].

Table 1: Details of the master audio tracks used during evaluation

Title	Artist	Genre
Sabattre	Endemic Species	Electronic
This Love	Madia	Jazz
Call me Crazy	Madia	Pop
Revolution	Chaos Kings	Rock

Table 2: System Parameter Values

Parameter	Value
K_d	0.12
Q_{PLL}	1/3
Cut-off frequency of preprocessing	1500 Hz
Constant gain factor x_{env}	0.95
Loop filter cut-off frequency f_C	20 Hz
Max $F0(N)$ values used	5
f_b	0.001
G (Equalisation's gain factor)	3 dB

3.1. Subjective Evaluation

During the listening tests, a total of eighty (80) subjects were selected to participate, most of them being students and the rest being members of the faculty staff of the department of Sound & Musical Instruments Technology, in the island of Kefallonia, Greece. All of the participants were familiar with audio mastering procedures, having successfully attended the courses of post-production and audio mastering provided by the above department. In addition to this, the members of the faculty staff are professionally involved with the mastering audio industry.

All the tests took place in the aforementioned department's professional audio studio and the complete set of equipment used is summarized in Table 3. During the subjective tests, the audio reproduction of both the original and enhanced audio data was performed through a stereophonic monitoring loudspeaker set.

Table 3: Equipment used for the subjective listening tests

Equipment	Brand & Model
Sound Level Meter	B&K 2250
Monitoring System	Klein & Hummel 0300D
Audio I/O	Digidesign 192 I/O
Controller	Digidesign C24
Software	Digidesign ProTools 7.3.1

For all four (4) unmastered audio tracks, no prior stage of dynamic processing was applied. Moreover, the enhancement of the proposed system lead to a slight level of increase at signal's RMS energy, compared to the original, unmastered audio material. For these two reasons, a calibration stage was introduced in order to prevent the biasing effect of different perceived loudness.

To do so, a stage of energy calculation for each track and calibration of the equipment was performed. During this stage, the available software's gain automation functions were used and every audio track was set to have an output of 85 dBSPL (with none weighting filter applied) at the location of the listening position, which represents a very frequently used sound level for monitoring purposes and interface calibration [11]. The sound pressure level was measured using the Type-A sound level meter appeared in Table 3 above. It should be also noted that the forte passages of the audio tracks were used for the measurement stage, while the divergence of the automation gain factor, among the different audio tracks, did not exceed the value of 1.3 dBFS.

After the calibration stage, the participants entered the studio one-by-one. Each one was informed about the experimental setup, it's objectives and the scoring options. Then, during the test, each subject heard both versions of the audio tracks, the unmastered and the enhanced output in a random order. Following the reproduction of the audio tracks, the subjects were asked to define the perceptually best audio track in terms of tonal balance enhancement for the aforementioned mastering processing. Finally, each participant's choice was tracked down, to create a preference percentage for each track's version (unmastered - enhanced).

3.2. Objective Evaluation

In order to estimate the audio output quality of the proposed system, a series of objective measurements were incorporated. Specifically, the Perceptual Evaluation of Audio Quality (PEAQ) standard for audio quality assessment was used. PEAQ was developed from International Telecommunications Union (ITU) and can provide comparison values between devices and systems, often used in the multimedia and high-end audio industry [8].

This measurement method models the fundamental properties of the human's auditory system by employing a difference measurement technique between a reference and a test signal [9] and resulting into audible difference values, also described as Subjective Difference Grades (SDG) [9], which are summarized in Table 4.

Table 4: Subjective Difference Grades

Imperceptible	−0.000
Perceptible, but not annoying	−1.000
Slightly annoying	−2.000
Annoying	−3.000
Very annoying	−4.000

For the purposes of the present assessment, a MATLAB implementation of the PEAQ assessment algorithm was used [10]. The reference signal that was employed as input to the aforementioned difference measurement was the unmastered/original audio, which was compared to the proposed system output.

4. RESULTS AND DISCUSSION

In this section, the results from both the subjective and objective experimental setups described previously are demonstrated. In particular, Table 4 summarizes the PEAQ-based difference grades for each track respectively, also including a mean value of all the measured grades. Clearly, from a range that spans from 0, meaning "Imperceptible", to -1, representing "Perceptible but not annoying", a mean grade of -0.754 has been achieved. Hence, it can be considered that the processed audio tracks achieve a percentage of transparency during the proposed enhancement. On the one hand, one can easily conclude that the proposed system does not perform well. However, on the other hand, in such kinds of processing procedures, significant and audible/perceptible changes in the spectral behavior may yield to unwanted results [1, 2]. As a consequence, the transparency often describes these kind of processes [1].

Table 5: Objective Evaluation Scores

PEAQ Grades	
Electronic	-0.403
Jazz	-0.555
Pop	-0.807
Rock	-0.860
Mean PEAQ Grade	-0.754

It can be also observed that between the Electronic/Jazz and Pop/Rock musical compositions, there is a divergence in the PEAQ score in the range of 0.35. The main reason for this observation is that the closely spaced used centre frequencies' (f_c) bandwidth, which were previously acquired from the histograms, affect the same spectral portion, thus providing a slightly larger boost in these regions. In addition to this, the difference between the processed frequency bands may significantly bias the PEAQ scores.

The results derived from the subjective experimental procedure are shown in Table 6. Clearly, a rate of 63/80 listeners actually preferred the automated tonal balance enhanced audio tracks. Considering their familiarity with mastering procedures and the usage of a professional studio as part of laboratory equipment for their studies, it can be inferred that the proposed system yields good results, providing an adaptive and robust option among the mastering processing chain.

Table 6: Scores Acquired from Listening Tests

Preference Rate Percentage	
Original Audio Preference Percentage	21.25%
Enhanced Audio Preference Percentage	78.75%

5. CONCLUSIONS AND FUTURE WORK

In the work at hand, a system providing automated tonal balance enhancement for audio mastering applications was presented. The system realization is based on an efficient pitch tracker that offers robust and fast sample-by-sample computations, combined with a set of second-order peaking-type filters. The proposed system performance was evaluated through PEAQ measurements and listening tests, achieving a mean PEAQ grade of -0.754 and a human preference rate in the range of 79%.

As it can been seen from the derived results, the proposed system has successfully achieved the aforementioned initial aims that motivated this work. In particular, during the subjective evaluation, the derived, tonal balance enhanced output obtained better scores among the majority of the experienced listeners that participated in the test sequence. In addition, the results of the objective PEAQ measurements show that the alterations introduced in the processed audio material do not affect it significantly in terms of human perception. Thus, the final output can be considered as a highly-similar copy of the original audio content, enhanced with perceptually efficient tonal balance characteristics.

Future implementations and improvements of the proposed approach may include the employment and assessment of alternative equalization strategies, as well as additional compression and spatial/overall enhancements, in a generalized attempt to minimize the required human involvement during audio mastering.

6. REFERENCES

[1] Owsinski B., "The Mastering Engineer's Handbook: The Audio Mastering Handbook," Second edition, Artistpro (2007 Dec.).

[2] Katz B., "Mastering Audio: The art and the science," Second edition, Focal Press, (2007 Oct.).

[3] Katz B., "The Secret of the Mastering Engineer," *in t.c. electronic* Articles (1999).

[4] Zolzer U., Sankarababu S., and Moller S., "PLL-Based Pitch Detection and Tracking for Audio Signals," *in Proc. Eighth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing* (Piraeus, Greece, 2012 Jul.).

[5] Zolzer U., "DAFX: Digital Audio Effects", Second Edition, John Wiley & Sons Press, 2011.

[6] Valimaki V., Gonzalez S., Kimmelma O., and Parviainen J., "Digital audio antiquing - Signal processing methods for imitating the sound quality of historical recordings," *J. Audio Eng. Soc*, vol. 56, no. 3, pp. 115-139 (2008 Mar.).

[7] Analog Dimension, (2012, Nov. 01). Mastering Audio Samples- before and after mastering. Available at: http://www.analogdimension.com/audiosamples.html

[8] Campbell D., Jones E., and Glavin M., "Audio quality assessment techniques - A review, and recent developments," *J. Elsevier Signal Processing*, vol. 89, no. 8 pp. 1489-1500 (2009 Aug.).

[9] Salovarda M., Bolkovac I., and Domitrovic H., "Estimating Perceptual Audio System Quality Using PEAQ Algorithm," *in Proc. 18th Int. Conf. on Applied Electromagnetics and Communications*, (Dubrovnik, Croatia, 2005 Oct.).

[10] McGill MMSP Lab, (2012, Oct. 14). Index of /Documents/Software/MATLAB/PEAQ. Available at: http://www-mmsp.ece.mcgill.ca/Documents/Software/Matlab/ PEAQ/

[11] Katz B., "An Integrated Approach to Metering, Monitoring, and Levelling Practices", Techn. Comm. *Audio Eng. Soc. on Studio Practices and Production*.