Audio Engineering Society

Convention Paper

Presented at the 127th Convention
2009 October 9–12 New York, NY, USA

Automatic equalization of multi-channel audio using cross-adaptive methods

Enrique Perez_Gonzalez[1], and Joshua Reiss[1]

[1] Centre for Digital Music, Queen Mary, University of London, London, E1 4NS, England
enrique.perez@elec.qmul.ac.uk, josh.reiss@elec.qmul.ac.uk

ABSTRACT

A method for automatically equalizing a multi-track mixture has been implemented. The method aims to achieve equal average perceptual loudness on all frequencies amongst all multi-track channels. The method uses accumulative spectral decomposition techniques together with cross-adaptive audio effects to achieve equalization. The method has applications in live and recorded audio mixing where the audio engineer would like to reduce set-up time, or as a tool for inexperienced users wishing to perform audio mixing. Results are reported which show how the frequency content of each channel is modified, and which demonstrate the ability of the automatic equalization method to achieve a well-balanced and equalized final mix.

1. INTRODUCTION

Equalizing a sound mixture is one of the most expert human tasks related to live music mixing. The main problem of determining the amount of equalization to be used is that the perceived amount of equalization is different from the physical amount of equalization applied. In order to achieve a perceptually pleasant equalization several things should be considered; whether or not the channel needs equalization at all, how many filters should be used, the type of filters and ultimately the amount of boost or cut they should have. Some studies on how the sound engineer performs these decisions have been performed by [1, 2].

Automatic mixing of speech and music levels has been attempted by [3-6]. However, very little has been done to attempt self-equalization of musical signals. The only notable example of such an attempt is [7], where the use of an off-line machine learning approach to the problem where humans need to manually train the machine. Once the machine is trained, it equalizes using nearest neighbor techniques. In this paper the authors will propose a method for use in live mixing situations driven by perceptual indicators. The proposed system does not require off-line machine learning methods. Instead it uses a real time cross-adaptive accumulative spectral decomposition approach to the problem based on a multiband implementation of [6]. A cross-adaptive algorithm is based on inter-channel dependency processing [8]. In other words, there are multiple inputs and outputs, and individual channel processing is dependent on the interaction with other channels. The cross-adaptive algorithm proposed uses the relationship between the perceptual loudness of all input channels to perform the equalization of every individual channel. The system then handles the task of weighting the channel equalization bands by using a perceptual indicator corresponding to a set of spectrally decomposed accumulation measurements of loudness. The spectral decomposition of signals is achieved by the use of a flat response filer bank. In this method we assume that the mixture in which loudness per band tends to the overall average loudness of the signal is a well-equalized mixture with optimal inter-channel equalization intelligibility. The idea behind this is to achieve an equal chance of masking between channels, thus optimizing the likelihood of each channel being heard. In order to achieve optimization, the system adapts its sub-band equalization gains according to the relationship of loudness indicators between channels and the overall average loudness. In this paper the authors will present the theory and implementation behind such a system, and results demonstrating the functionality of the system.

2. CROSS-ANALYSIS

The proposed system consists of two fundamental parts. First is the signal processing side of the algorithm, consisting of an equalizer. In the context of the presented algorithm, the equalizer under study has fixed frequency bands and the only parameters are the gains of each frequency band. The processing side takes the input signal channels $f_m(n)$ where m correspond to the channel number from $m=1$ to M, and outputs an equalized version of each channel known as $feq_m(n)$. The second part is the cross-analysis, which takes as an input $f_m(n)$ and outputs the control gain parameters corresponding to the equalization bands of each channel, $G_{km}(n)$, where K corresponds to the maximum number of equalization bands per channel and has a valid range $k=1$ to K. The system diagram of the overall system is presented in Figure1. It is the aim of the following sections to explain the theory and implementation of all the analyses stages required to derive the equalization parameters $G_{km}(n)$.

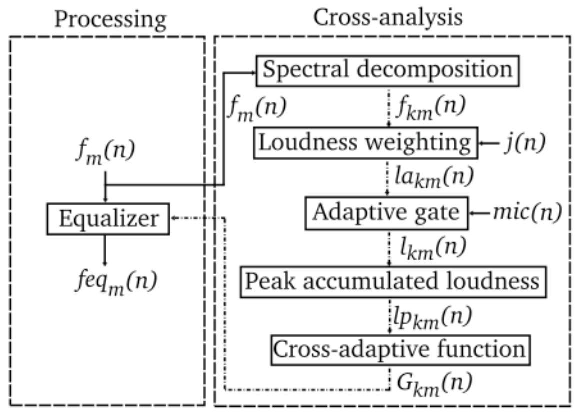

Figure 1 System overview block diagram

2.1. Spectral Decomposition

The first step of the cross-analysis is the decomposition of each of the channel inputs, $f_m(n)$, into K frequency bands. For each $k=0$ to K decomposition band there must be a corresponding equalizer band in the processing side of the algorithm. For this $f_m(n)$ must be processed by a filter bank, a fast Fourier transform, FFT, or a similar transform such as a constant Q transform, in order to separate the input into K spectral bands. For the system to perform properly the spectral bands or bins must be spread as evenly as possible and must add to a flat frequency response. The accuracy of the final performance of the system will be dependent on the amount of spectral decomposition bands; the more spectral bands used, the more accurate the system will be.

2.2. Loudness weighting

A model containing the ISO 226 standard loudness curves [9] has been used. Given that a psychoacoustic model has been used for weighting the signals, the more accurate the psychoacoustic model of loudness the better results the system should give. The model proposed, depicted in figure 2, consists of a look-up table containing all the coefficients necessary for generating the loudness weighting curves, $w(j(n))$. The table is driven by a sound pressure level, SPL, measurement denoted as $j(n)$. The ISO curves are defined in intervals of 10dB. The SPL value of $j(n)$ can be time varying, as obtained from a single SPL measurement microphone at the mixing position, or can be non-time varying by fixing the value of $j(n)$ for all values of n. Given that the spectrally decomposed input, $f_{km}(n)$ is weighted by $w(j(n))$ we can perform an averaging of length *s* in order to include in the model a way of modeling longer and shorter term loudness measures. Therefore the psychoacoustic weighting is the given by:

$$la_{km}(n) = \sum_{n=1}^{s} f_{km}(n) w(j(n)) / s \qquad (1.)$$

Where $la_{km}(n)$ represents the psychoacousticaly weighed signal derived from the spectrally decomposed signals $f_{km}(n)$.

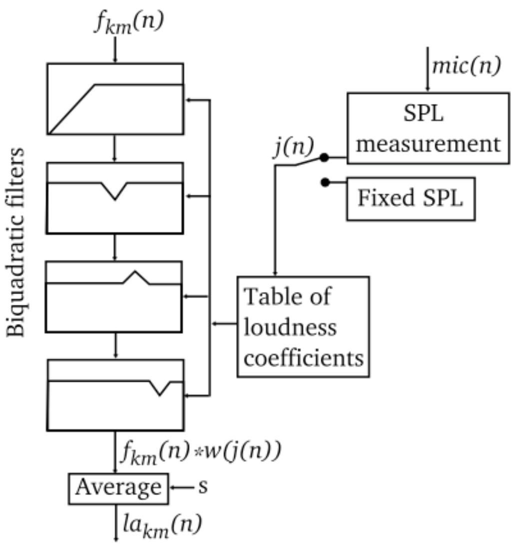

Figure 2 Loudness feature weighting diagram.

2.3. Adaptive Gating

The adaptive gating used in the proposed method is a multiband implementation based on an adaptive gating method implemented in [5, 10]. The main idea is to have a means of adaptively validate the data coming out of the loudness weighting stage, $la_{km}(n)$, in order to remove any noise. Cleaning $la_{km}(n)$ from noise will render a better spectral band loudness estimation. The principle is that for a given microphone there is a usable rage where it will capture the sound source in such a manner that the source signal is larger than the ambient noise. Therefore from the point of view of the microphone anything outside this usable range is considered noise. We can use a single microphone, *mic(n)*, such as the one used for measuring the SPL in the previous step, and place it outside of the usable stage microphone area, for example next to the mixing board. Then by determining the relationship between the ambience noise microphone and the individual inputs it is possible to determine if a signal contains noise or not. Such functionality is given by the following pseudo-code:

$t_k(n)=$ *[if $la_{km}(n)<=lamic_k(n)$,*
then open_gate,
else close_gate]

Where $t_k(n)$ is the adaptive threshold signal for operating the noise gate. $la_{km}(n)$ corresponds to the average loudness weighted signal of the spectrally decomposed input channels, given that $lamic_k(n)$ corresponds to the average loudness weighted signal of the spectrally decomposed ambiance microphone input. Therefore the loudness weighting signal, $la_{km}(n)$, can be gated by the adaptive threshold signal $t_k(n)$ in order to obtain a noise-free signal which is representative of the perceived loudness, $l_{km}(n)$. Such an implementation is depicted next:

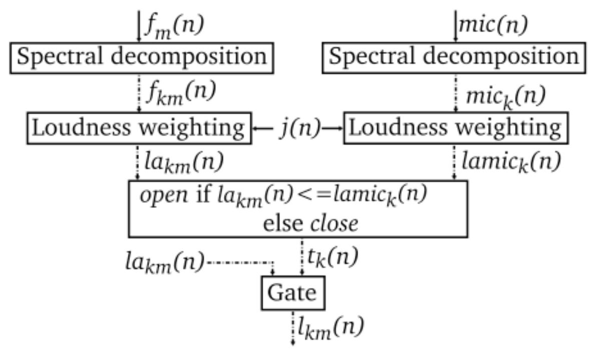

Figure 3 Multiband adaptive gate diagram.

2.4. Peak Loudness Accumulation

Ones we are sure that we have a noise-free signal which is representative of the perceived loudness, $L_{km}(n)$, we can proceed to determine its accumulated peak loudness. The proposed method for obtaining a value representative of the spectral band is to accumulate its normalized histogram in order to determine the mass probability function of the analyzed loudness band. From this probability mass function we can then determine the most probable loudness value for a given spectral band. Given the on line use of the algorithm, it is necessary to ensure, that the histogram variance is kept within range. This is due to the fact that knowning the maximum peak value of $f_m(n)$ does not ensure that the limits of the histogram values will be the same. This is because the peak magnitude of $f_m(n)$ is not the same as the weighted $la_{km}(n)$ peak value due to the loudness weighting. For this reason a cross-rescaling mechanism has been implemented. The system works by rescaling $l_{km}(n)$ by a factor $r(n)$. The overall system gain reference, $r(n)$, is given by finding the maximum gain value that can satisfy all $l_{km}(n)r(n)$ such that its maximum peak value is equal to one. This is accomplished by the following pseudo-code:

$rs_{km}(n)$= *if* $|Binmax_{km}(n)|>0$,
then $rs_{km}(n)=rs_{km}(n-1)-d$,
else $rs_{km}(n)=rs_{km}(n-1)$

Where $Binmax_{km}(n)$corresponds to the value taken by the highest bin of the histogram and $rs_{km}(n)$ corresponds to the maximum channel gain such that $max(l_{km}(n)\ rs_{km}(n))<=1$ so that the channel has a $Binmax_{km}(n)=0$. Where $r(n)$ is given by the following equation:

$$r(n)=\min(rs_{km}(n)) \quad (2.)$$

Where $r(n)$ is simply the minimum value over all $rs_{km}(n)$. Given that the scaling is within range, now we can proceed to look for the mass probability function peak, $lp_{km}(n)$ which should correspond to the most probable loudness value taken by a given channel band. The flow diagram of such a histogram rescaling system is depicted next:

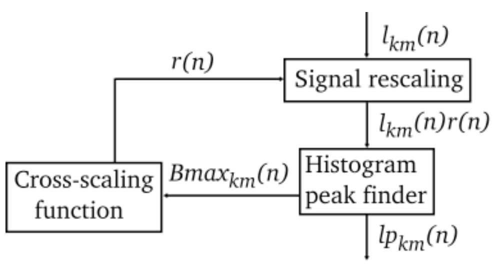

Figure 4 Peak loudness accumulation diagram.

2.5. Cross-adaptive function

The final signal processing of the equalized channel signals, $feq_m(n)$ for a set of channel inputs, $f_m(n)$, has the following equation:

$$feq_m(n)=EQ[f_m(n),G_{km}(n)] \quad (3.)$$

Where $G_{km}(n)$ corresponds to the equalizer band gain coefficients. Then, in order to achieve a continuous variation of $Gkm(n)$, so that a common average loudness between all channels and their corresponding equalization bands is achieved, the system can be modeled with the following system diagram:

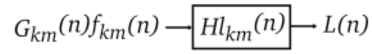

Figure 5 Loudness feature diagram

Where figure 5 is a multiband extension of the cross-adaptive model presented in [6], and from its transfer function the corresponding equation for determining the equalizer band gain coefficients, $G_{km}(n)$, can be derived as follows:

$$G_{km}(n)=\frac{L(n)}{Hl_{km}(n)f_{km}(n)} \quad (4.)$$

Where the numerator $L(n)$ is the average loudness of all channel equalization bands. Therefore $L(n)$ is given by the following equation:

$$L(n)=\sum_{m=1}^{M}\left[\sum_{k=1}^{K} lp_{km}(n)/K\right]/M \quad (5.)$$

Where $lp_{km}(n)$ corresponds to the most probable loudness value for each spectral band. M corresponds to the maximum number of channels involved in the mix and K corresponds to the number of bands in the spectral decomposition.

Then we can say that $Hl_{km}(n)$ is the transfer function of the combined block diagram of figure 3 to 5. Which is given by $Hl_{km}(n)=lp_{km}(n)/f_{km}(n)$. So that $lp_{km}(n)=Hl_{km}(n)\ f_{km}(n)$, where $lp_{km}(n)$ corresponds to the most probable loudness state per spectral band, and is therefore the denominator of equation 3.

In order to maintain system stability a system that will maintain maximum gain before feedback has been implemented. The system avoids the use of transfer function gains above unity in order to maintain system stability [11].

3. IMPLEMENTATION RESULTS

A five filter, first order, filter bank with flat frequency response was implemented in order to test the proposed system. The spectral decomposition filter bank consisted of the following Butterworth designs: a low-pass filter with a cut of frequency of 63Hz, three band-pass filters with mid band frequencies at 127Hz, 750Hz and 4000Hz and a high-pass filter with a cut off frequency of 8000Hz. Such implementation has been depicted in the top section of figure 6. Its corresponding equalizer design makes use of the same filter topology, as shown in the bottom plot figure 6.

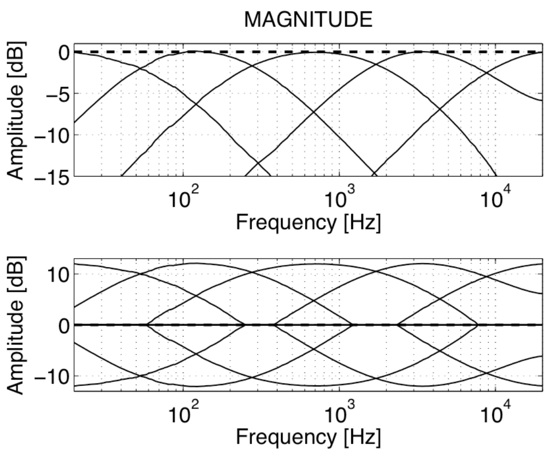

Figure 6 Top, filter bank transfer function with K=5. Bottom, corresponding equalizer transfer function with K sub-bands. Individual filter response has been plotted as a solid line; combined frequency response at unity gain position has been plotted as a dashed line.

3.1. Music Signals

A set of eight channel, live multi-track recordings with different styles of music was used for testing the system. A single omni-directional flat frequency response microphone was used to capture the ambient and calculate the SPL. Clear spectral changes were heard after applying the algorithm. First impressions are positive but no proper subjective evaluation was carried out. The system seems to perform better in the high frequency range than in the low end. Spectrum comparisons indicated a tendency to increment the high frequency range. An increase of dynamic range on the auto-equalized signal was encountered in all signals tested. For all the signals tested a increase of 3dB crest factor was observed. This can clearly be seen in figure 7, where the un-equalized time domain signal seems to present more cluttering, while the auto-equalized signal has more defined transients. The recording used to generate figure 7 had an increase of crest factor on the order of 3.2 dB.

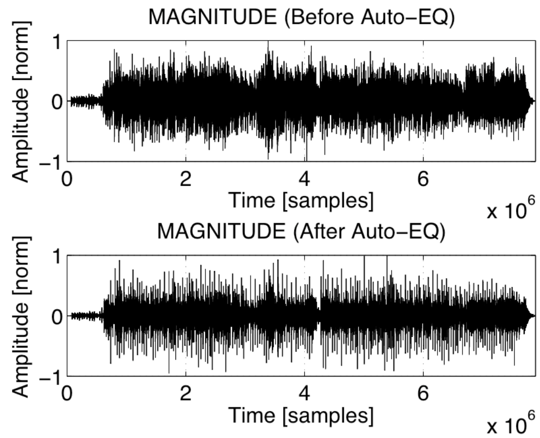

Figure 7 Time domain self equalization of a music signal

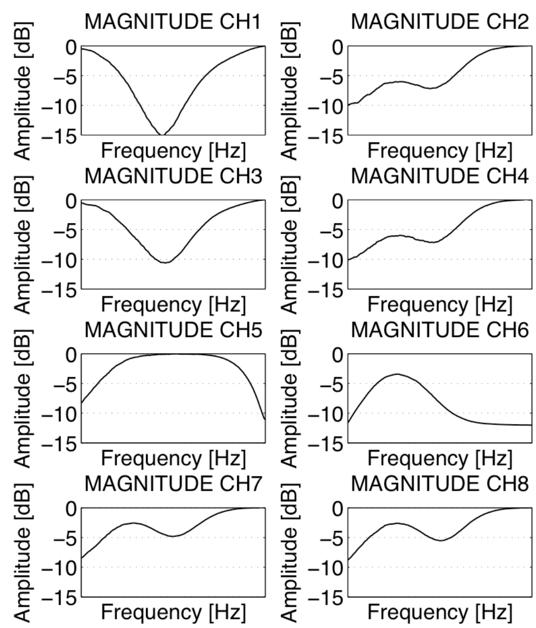

Figure 8 self-equalization of a music signal

The resulting equalized transfer functions for an auto-equalized track recording consisting of Ch1= vocals, Ch2= guitar, Ch3= synthesizer-left, Ch4= synthesizer-right, Ch5= Snare, Ch6= kick-drum Ch7= high-hat and Ch8= over-head, were plotted in figure 8. In all samples tested it was clear that the algorithm tended to improve the high frequency section of the spectrum but seems to have a tendency to under boost low frequencies.

3.2. Test Signals

In order to find out if this low frequency under boosting was due to the algorithm itself or the implementation, a test signal measurement consisting of white noise was input to the system. If the system was performing as expected it should match the inverted loudness weighing curve applied to the signal, 1/ *w(j(n))*. It was found that for high SPL levels the system performs as expected, see top plot figure 8, where the equalizer transfer function gives a good match to 1/*w(120dB)*. On the other hand it fails to match the low frequency spectrum section when below 1/*w(90dB)*, bottom plot figure 8. It was found that below 90dB the selection of having a LPF with a cut-off frequency of 63Hz caused cussed the analysis to have no low frequency signal available in that lower sub band. This means that for being able to approximate the white noise signal to a loudness curve of 1/ *w(j(n))* for *j(n)<=90dB* it is necessary to have an analysis bank with more than five filters or a higher cut-off up point for the LPF. Current research efforts concentrate on achieving the optimal spectral decomposition analysis method to maximize performance.

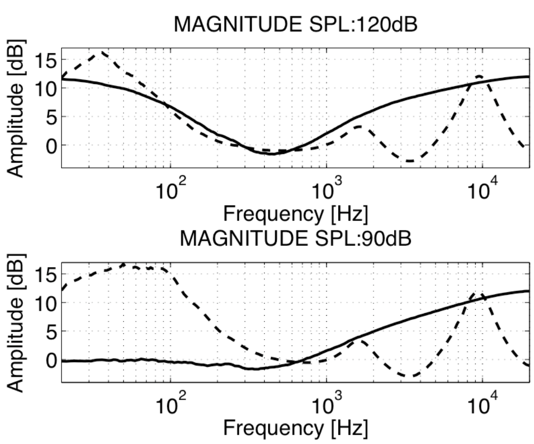

Figure 9 Self-equalization of a white noise test signal, solid line. Top, auto-equalized response for *j(n)=120dB*. Bottom, auto-equalized response for *j(n)=90dB*. Bottom. 1/ *w(j(n))* is represented by the dashed line.

Results indicate that a better implementation either with more filters in the filter bank or a Fourier approach will greatly improve the low frequency implementation. A formal subjective evaluation must be performed to evaluate the algorithm and to determine its performance against a human audio engineer.

4. CONCLUSIONS

The theory and implementation of a cross-adaptive system capable of using perceptual weighing in order to achieve equal probabilistic psychoacoustic weighing of the equalizer bands has been presented. Current implementations indicate the system has potential uses in live equalization for music. Results produced using a five-band spectral decomposition implementation indicate that a Fourier based spectral decomposition, together with a corresponding Fourier based equalizer could dramatically improve results. More system profiling in order to evaluate the subjective performance of the system is needed.

5. REFERENCES

[1] J. Bitzer, and et al, "Evaluating perception of salient frequencies: Do mixing engineers hear the same thing?," in *124th AES Convention*, 2008.

[2] J. Bitzer, and J. LeBeuf, "Automatic detection of salient frequencies.," in *126th AES Convention*, 2009.

[3] E. Campbell, and Whittemore R. T., "Automatic microphone mixing apparatus " in *European Patent Office, US 4357492 (A)* 1982.

[4] R. B. Dannenberg, "An Intelligent Multi-Track Audio Editor," in *ICMC*, 2007, pp. 89 - 94.

[5] D. Dugan, "Automatic Microphone Mixing," in *AES*, 1975.

[6] E. Perez_Gonzalez, and J. Reiss, "Automatic Gain and Fader Control For Live Mixing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* NY, 2009.

[7] D. Reed, "A perceptual assistant to do sound equalization," in *5th international conference on Intelligent user interfaces* New Orleans, Louisiana, United States, 2000.

[8] V. Verfille, and et al., "Adaptive Digital Audio Effects (A-DAFx): A New Class of Sound Transformations," *IEEE Transactions On Audio, Speech, and Language Processing,* vol. 1558-7916, 2006.

[9] ISO, "Acoustics — Normal equal-loudness-level contours, ISO 226:2003(E) ", 2003.

[10] D. Dugan, "Application of Automatic Mixing Techniques to Audio Consoles," in *AES*, 1989.

[11] E. Perez_Gonzalez, and J. Reiss, "An automatic gain normalisation technique with applications to audio mixing," in *AES*, 2008.