iMaster - Automated Mastering of Digital Audio

1. Problem Description

Mastering is the process of making a final mix down of a song or track that has already been produced. It can involve the adjustments of levels, equalization, compression and limiting which provides consistency across songs and formatting to CD quality, with a sampling rate of 44100 at 16 bit, etc.

Mastering of audio signals can be a laborious and time consuming task that requires years of experience and expertise. Many musicians have little resources and money to spend on getting a mastering engineer to finalize their productions and generally have little knowledge and expertise on how to deliver a high quality final mix. Good mastering can make a track sound spacious, have width and image, where individual instruments can breathe. Typical bedroom mix downs lose their sonic attributes once they have left the confines of the room and are taken into a different acoustic environment. Many producers find it hard to master their own tracks and often try to implement processing which fails to give the same sonic appreciation felt in professionally mastered songs.

Using the new iMaster plug-in from iMDeafAudio, producers can master and enhance their mixes with a click of a few buttons. The mastering plugin has a host of features. However, what sets iMaster apart from the crowd, is its high end Harmonic Exciter algorithms.

The iMaster features several different harmonic exciters which can be implemented and altered to suite the final mix-down. These include: Vintage and Modern.

Harmonics occur when the natural frequency of a waveform resonates in a periodic nature, this is typically what makes musical instruments sound pleasant. The waveform features a fundamental frequency i.e. the lowest frequency produced by the waveform, which is then followed by a series of integer multiples of this frequency (Smith, 1997). The harmonics archetypally decrease in level as frequency increases, however the level of the individual harmonics can vary at specific periods. For example a clarinet has harmonics with increased level at the odd harmonic integers compared with the even.

2. Specification

The harmonic excitation implemented by iMaster consists of, dynamic equalisation, phase manipulation, harmonic synthesis and harmonic distortion.

2.1 Spectral Analysis

The Harmonic synthesizer adds high-frequency synthesis for reconstruction of signal details buried in noise. It creates harmonics of selected frequency bands of the audio being processed, and mixes those harmonics in with the original audio. Increasing Synthesis can increase the sense of life and clarity in processed audio. Too much Synthesis may cause apparent distortion in the signal.

The ‘Modern’ algorithm works from Fourier’s analysis and synthesis (Risset, 1985). Fourier discovered that any periodic waveform can be constructed by adding harmonic sin or cosine waves to a relative sine wave.

Once a Fourier analysis is performed on the musical content to ascertain its harmonic frequencies, a re-synthesis algorithm can be applied to re-insert missing or weak harmonics through the creation of a waveform from sinusoidal oscillators with varying amplitude and phase.
There have been many different approaches to the harmonic synthesis since the 1970’s, most notably by Moorer (1978) who implemented the ‘Phase Vocoder’ as a means of resynthesizing sounds. The Phase Vocoder performs both analysis and synthesis of the signal and determines the harmonic representations off the waveform at the output. The signal is modelled from a sum of sin waves that are time-varying in amplitude, phase and frequency. It can potentially be used for a variety of sources that are not periodic in nature such as speech. However, it relies on predicting the waveform for re-synthesis and thus struggles to replicate more percussive instruments such as pops and clicks that have high transients.

[Fig 1. Filter bank interpretation of the Phase Vocoder. Own Diagram; Source (Dolson, 1986)]

As you can see from (fig.1) the vocoder re-synthesis algorithm has a fixed bank of band pass filters. The filters act to separate the audio signal into segments so that the individual amplitude and frequency components can be extracted from them, which are then used in recreating the sinewaves at the oscillators. In terms of harmonic re-synthesis there needs to be one filter per harmonic frequency that is to be resynthesized. This method can become computationally heavy, as for a fundamental frequency of 250 Hz, there are harmonics at every 250 Hz beyond that, up to the human hearing threshold of 20,000Hz, is equal to 100 filter banks.

\[
f/20,000 = \text{Filter Banks}
\]

For non-harmonic and polyphonic sounds this becomes a greater challenge as the overtones are unevenly spaced and thus are greater in number (Dolson, 1986).

2.2 Short-time Fourier Transform Technique

An improved method of the Filter bank method is the Fourier-Transform Interpretation which is implemented in the iMaster plugin. This method uses filter bins, which is in essence, the same concept as filter banks, whereby the waveform is converted into the frequency domain, which is divided into bins, representing individual frequency bands. The transform focuses on the individual frequency and amplitude values of the different bins. The Fourier transform is a simplification of the Fourier series, which analysis the amplitudes and phase components of sine waves that must be added together to create a complex periodic waveform.
The Short Time Fourier Transform (STFT) and a spectrogram are used to analysis signals with varying frequency with time. The STFT is used in cohesion with a windowing functioning. The windowing function acts as the band filter, whereby it analysis a section of time and slides the data through the window. The STFT equation is given below:

\[ X[n, k] = \sum_{m=0}^{L-1} x[n + m]w[m]e^{-j\frac{2\pi}{N}km} \]

Source: (Desainte et al. 2000)

The iMaster algorithm procedure tracks the peaks within a waveform by using a STFT, looking for harmonic tones, it then extracts the sinusoidal peaks, which are then removed by spectral subtraction and a process of additive synthesis is then applied to parts of the spectrum that are lacking in depth.

3. **Fourier Synthesis**

Fourier synthesis is a mathematical equation for deriving a specific waveform from combining a sine wave signal with sine or cosine wave harmonics related to the fundamental frequency of the initial sine wave. The assembly or synthesis of a wave is done discreetly by adding the instantaneous amplitudes of the products of the sin and cosine waveforms. It is given by the equation:

\[ v(t) = a_0 + \sum_{n=1}^{\infty} \left[ a_n \cos(n\omega_1 t) + b_n \sin(n\omega_1 t) \right] \]

Source: (Kahrs, 2002)

4. **Harmonic Distortion**

While additive synthesis is a process of recreating harmonics from sin and cosine wave, which result in a cleaner sound, harmonic distortion uses non-linear distortion to create additional harmonics which only affects frequency above the fundamental. It is mainly used for higher frequencies by the introduction of a high pass filter. It is used to add brightness and clarity to the track with a warmth and fullness that is appreciative from the analogue era. It uses a process of non-linear distortion which is added through the subtle clipping of a waveform. This can be given by the simple expression:
\[ y = a_2 x \]

Whereby \( y \) is the output signal and \( x \) is the input, multiplied by \( a \), the gain factor (Chalupper, 2000).

When a waveform is clipped and altered into a square wave shape it produces frequency modulations of its fundamental frequency, which result in increased harmonics.

The iMaster implements a 2\(^{\text{nd}}\) Order Butterworth high-pass filter which gives a flat frequency response in the passband i.e. no ripple, followed by an asymmetric soft clipper. Soft clipping ensures only low-order harmonics are generated, by clipping asymmetrically, both odd and even harmonics are generated (Shekar et al. 2013). The signal chain for this process is outlined below. You can see that it operates in two chains, one for the original signal and one for the output, which are then combined.

![Own Diagram; Source (Shekar et al. 2013)](image)

The equation that defines the Butterworth filters frequency response is given by:

\[
|H(j\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_0}\right)^{2N}}
\]

Source: (Mello et al. 2007)

5. **Dynamic Equalisation**

A dynamic equaliser can react to different thresholds of sounds with an attack and release time, similar to that of a compressor. This is useful in mastering applications when a buildup of sounds affect how certain instruments perform, allowing an equaliser at certain thresholds, maintains a coherence between lower and louder levels.

6. **Implementation**

The basic goal of the Harmonic Exciter is to boast and provide exaggerated content to a specific band of frequency, enhancing the sonic attributes of the song. The application outlined above can operate on
multiple bands of frequency at one time and features adjustments for both additional harmonic content and amplification. This can be utilized on individual instruments or on whole tracks for mastering.

It will be implemented as a digital audio workstation plugin and have several adjustable features such as the amount and type of processing that can be added i.e. vintage or modern, the frequency band of interest and parametric dynamic equalisation on each band.

Depending on the application, it can be used for enhancing vocal content, or providing extra dynamic harmonics to the bass or high end of the track and as a preset to mastering the overall sound of a song.

The ‘Modern’ algorithm will use additive synthesis whilst the ‘Vintage’ will use harmonic distortion, due to the nature of the process and the sound it produces. The additive synthesis produces a cleaner unnatural sound often found in electronic music whilst the harmonic distortion produces a fuzzy warm effect, often associated with an analogue system.

7. Evaluation

Once the plugin has been designed and programmed it is evaluated to see if it meets the engineering specification outlined in the brief for designing the product. This consisted of a series of objective measurements. These measurements provided a reliable benchmark for assessing the performance outcome of the plugin and will completely quantify the performance of the device (Ballou, 2005).

7.1 Frequency Response

The frequency response of the Input and output waveform is evaluated and compared using a Fast Fourier Transform of the time domain signal. This provided data on how well the plugin reproduces harmonics and frequency content, but also maintains the original data and representation of the signal. To perform this test an impulse with a flat frequency response was processed using each algorithm independently. This allowed for any changes in frequency content to be observed very easily (Shekar et al. 2013).

7.2 Sine Wave Analyses

A sine wave at varying frequency was processed through the plugin to analyse the widening of the spectrum by the exciter. It clearly demonstrates a widening of the spectrum and the enhancement of the original content at specified frequency bands (Shekar et al. 2013).

7.3 Speech Intelligibility Index

One of the key factors in the performance of a harmonic exciter is how clear and bright it can make the audio sound. A speech intelligibility test is a good method for testing how well the exciter can introduce additional information i.e. harmonics, which can enhance the clarity of the audio, whether it be for speech or music. Two tests were performed one with harmonic excitation turned on and one with it off. For this test it has been shown that a harmonic exciter does increase STI results, but will only perform well with a high signal to noise ratio on the given vocal sample (Chalupper, 2000).

Overall the plugin has been specified to perform an audio enhancement function for audio production and mastering assistance, it performs well at the given tests and conforms to the engineer’s specification.

It is impossible for objective measurements to foretell whether the desired audio processing is going to be pleasant to ones ear, it is therefore desirable to correlate subjective impressions of listeners with object design parameters to maximize the effectiveness of the evaluation (Ballou, 2005).
5. Perceptual Assessment

To assess whether the iMaster plugin lives up to the foresaid hype, it is proposed that a blind ABX listening test will be performed with audience members from the general public. Whereby, the subjects will not be told whether track A or B is the processed track and subsequently, they will have to determine which one it is i.e. X (Nousaine, 1991). Using a blind test alleviates any unconscious bias towards the processed track. Listening is a very subjective nature and can vary greatly between individuals, therefore through using the general public rather than trained critical listeners it will provide a simplified approach to the listening evaluation, rather than a more in-depth and critical approach. The idea is to ascertain whether people can see performance enhancements within the harmonically excited track over the non-processed.

This approach is subject to preferential differences between each listener, perhaps one listener actually prefers the non-processed signal over the processed (Clark, 1981). This leads to a listening test design which incorporates a selective demographic who’s preferred listening genre is popular music, this will be achieved through post-screening. Popular music is one of the genres of music which particular lends itself to using such techniques for audio production and therefore listeners will be more inclined to its nature.

Another fact of the test will be the style in which the questions are asked, simply asking whether they prefer one track over another would not be sufficient. Instead a series of questions will be asked which relate to the tracks sonic attributes (Bech & Zacharov, 2006).

1.) Is track A perceptually different from track B?
2.) Which track sounds like it is more proficiently produced?
3.) Which track sounds more spacious and bright?
4.) Which track do you prefer?

Using questions in this nature will help alleviate bias from individual preferences and produce results which are accurate in determining which track is sonically more appreciable.

8. References


Glen Ballou, 2008 Handbook Sound Engineers, Taylor and Francis, Kentucky, USA


