Automatic target mixing using genetic optimization of gain and equalization settings.

Daniele Barchiesi and Josh Reiss. Centre for Digital Music, Queen Mary, University of London.

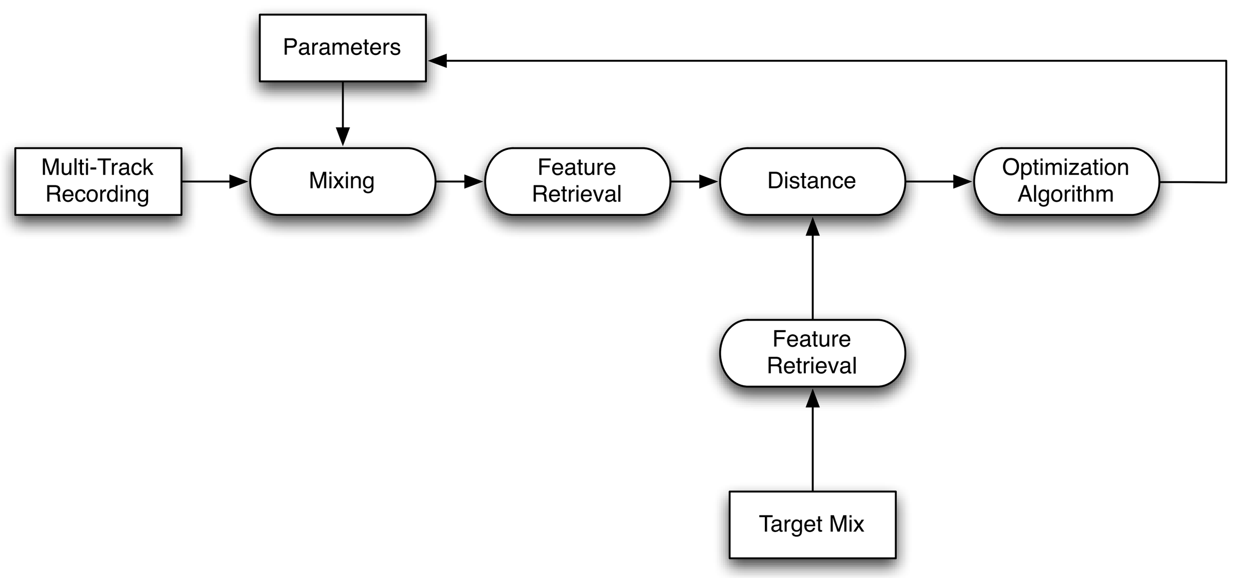
What is the Automatic Target Mixing?

The goal of automatic target mixing is to build a mix from a multi-track recording which is “similar” to a target mix that the user can choose as a representative of a musical genre or as an artistic reference.

The mixing process usually requires a skilled sound engineer who adjusts many parameters like gains, equalization settings and other effects for each track of the recording session.
With the development of this technique, even a non expert user could obtain mixes that “sound like” the professional ones. Moreover, an automatic target mixing can be a good starting point for the mixes made by the expert users.

The proposed technique has been applied to adjust the gains of a multi-track recording. (Bennett Kolasinski, A framework for automatic mixing using timbral similarity measures and genetic optimization, Proceedings of the 124th Convention of the AES, May 17-20 2008).
The aim of this work is to improve the robustness of that algorithm and to extend the automatic target mixing framework to the equalization of the multi-track recording.

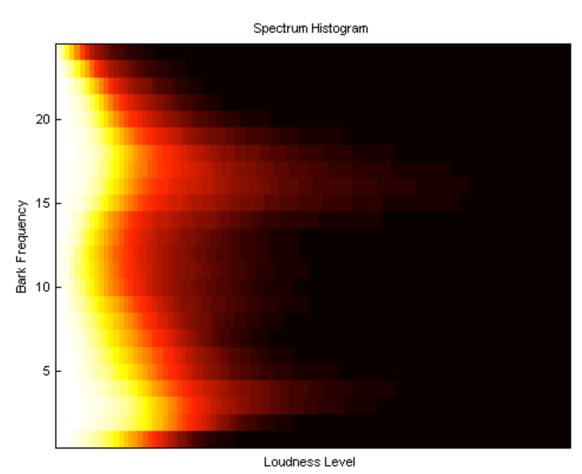
The Automatic Target Mixing Framework.



There are three key concepts in the automatic target mixing technique:

- **Parameters** are the variables which can be adjusted to make the mix from the multi-track recording.
- **Metric** is the distance between the features retrieved from mix and target. It is used to compute the similarity of the two mixes and is the cost function of the optimization algorithm.
- **Optimization Algorithm** is used to find the parameters set that minimize the metric, that is the set that produces a similar mix.

Spectrum Histogram



The Spectrum Histogram of a signal depicts how many times a certain loudness level is reached or exceeded in each of the 24 bands of the Bark frequency scale.
The euclidean distance between the SH of the mix and the target is the metric chosen for the automatic target mixing algorithm.

Genetic Optimization

The SH is a function of the mixing parameters. The GO searches the parameters' space in order to find the global minimum of the metric.
The GO starts with a random population of M individuals, which are points in a n-dimensional parameters' space (where n is the number of tracks to be mixed), and applies three operators in order to find the best parameters set.

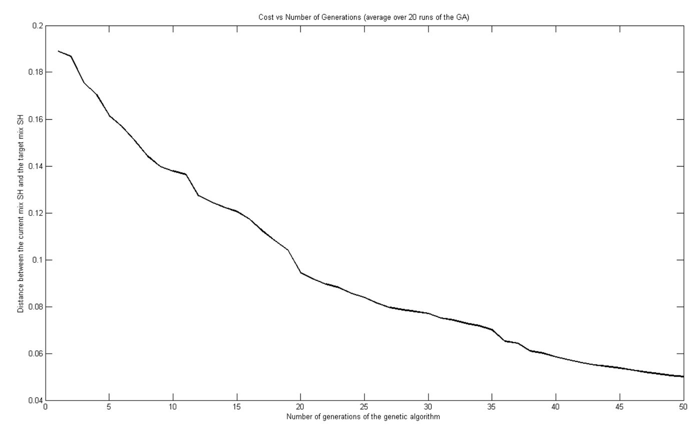
- **Selection** selects the most fit individuals, that is the individuals that generate a mix whose SH is close to the target SH.
- **Crossover** combines the information contained in pair of the selected individuals and generate a child from them.
- **Mutation** randomly alter one parameter of the selected individuals.

The GO runs until a fixed number of generations is reached and returns the best found parameters set.

The Gain Problem.

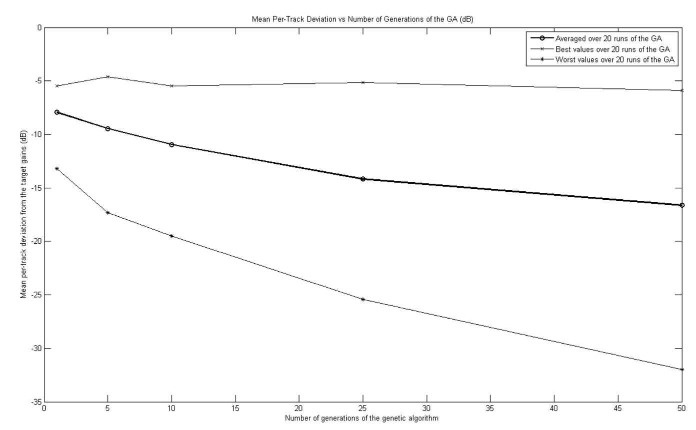
In order to test the automatic target mixing framework, a prototype target mix is made from a multi-track recording applying random gains to each track. Then the same audio material is the input of the target mixing algorithm, which is supposed to find the random gains applied to the target.

Evolution of the cost function of the GO for a 8-tracks recording.



The plot depicts the value of the metric for the best individual in each generation of the GO.
The number of individuals in each population is 25 and the maximum number of generations is 50.

Mean per-track deviation from the target gains.

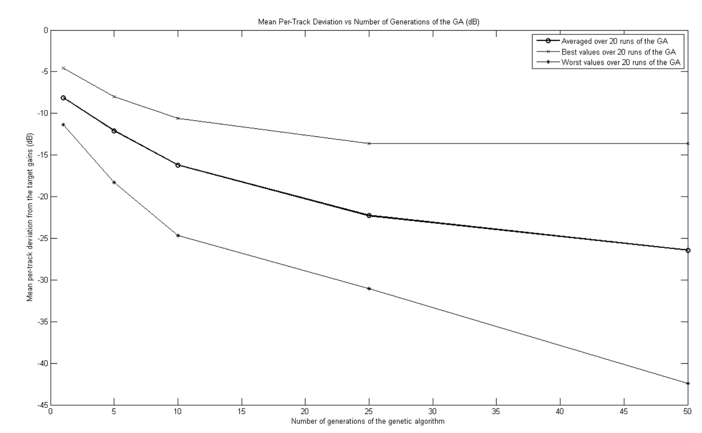


The plot depicts the mean deviation from the target gains of the best solutions found at the 1st, 5th, 10th, 25th and 50th generation of the GO.

The test has been run for 20 times, and the right gains set has been found in the 80% of the runs.
In the remaining 20% the algorithm hasn't converged to the right solution.

Robustness of the Genetic Optimization

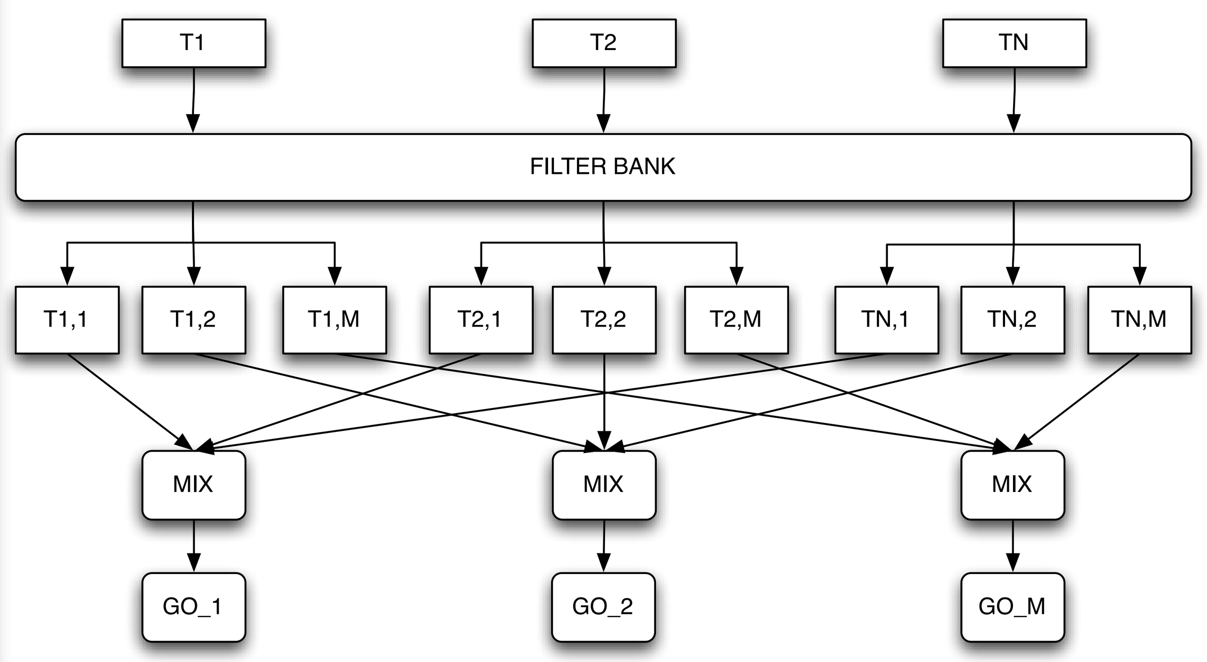
In order to improve the robustness of the algorithm, several variables of the GO can be adjusted. The following plot depicts the mean per-track deviation from the target gains using GO with a population of 50 individuals instead of 25.



Although 5% of the algorithm runs hasn't converged to the right solution, the deviation from the target gains shows a general improvement with a higher population size.

Other tests has been done increasing the rate of mutation and using a variable population, that is a population which has a greater number of individuals in the first generations and few individuals in the last generations.
In all cases the performances of the target mixing algorithm improved, at the expense of greater computational times.

The Equalization Problem: a simple approach.



Each one of the *N* tracks is divided in *M* bands by the filter bank. Then signals belonging to the same frequency band are mixed and the GO is performed for M independent gain problems.

This approach can be viewed as an M-bands graphic equalizer, therefore the target mixing can be effective detecting the settings of this type of equalization.

The general-case equalization problem: finding the right metric.

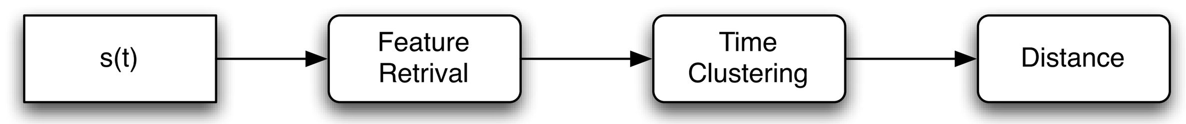
When thinking about real-life applications of the automatic target mixing technique, we have to deal with different audio material for the mix and the target. Moreover, considering the equalization problem, we can't have prior knowledge about which type of equalizer has been applied to each track.
The choice of the euclidean distance between spectrum histograms doesn't guarantee that minimizing this kind of metric corresponds to finding the right equalization settings.

In order to find the right metric for the EQ problem, four altered versions of a test song has been produced, and several different metrics has been tested on these signals.
The altered versions of the test song are:

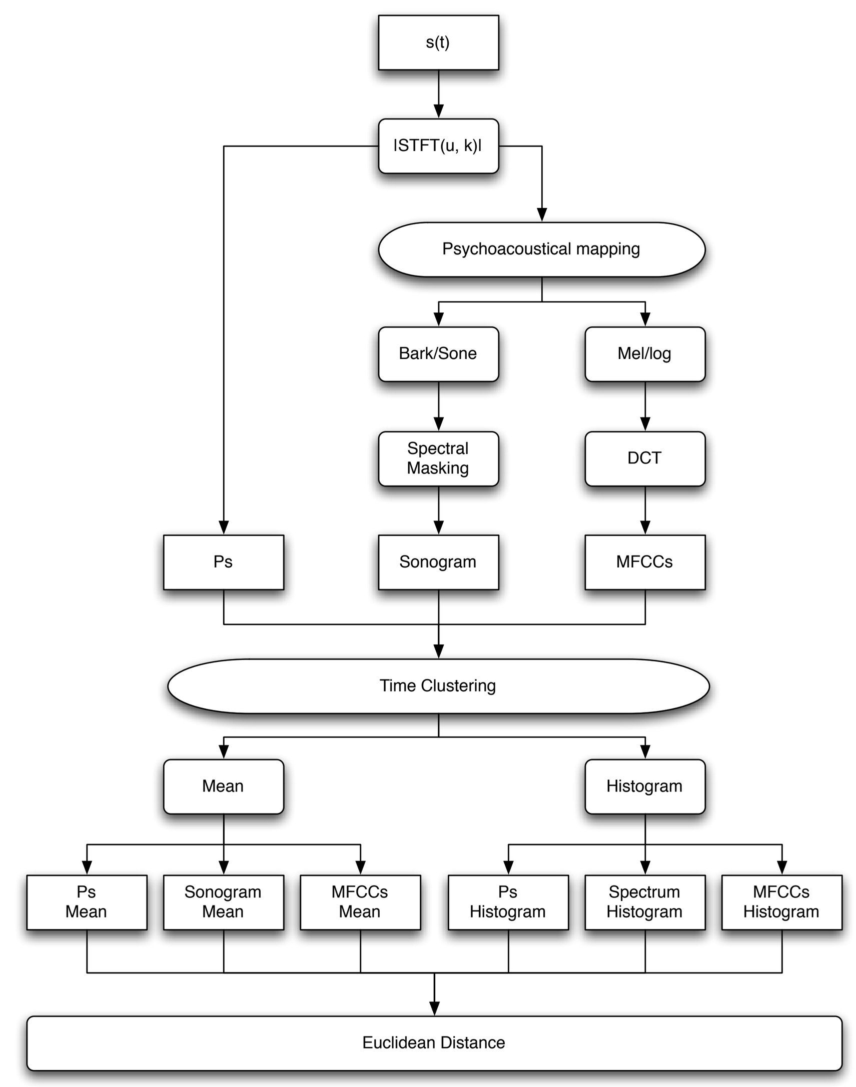
- **EQ10**: each track has been equalized with a parametric equalizer applying a total of 10dB equalization.
- **EQ30**: each track has been equalized with a parametric equalizer applying a total of 30dB equalization.
- **M1**: the modified1 song is obtained through the re-arranging of the test song's audio material.
- **M2**: the modified2 song contains different audio material than the test song, but share with it the same number and kind of instruments, the same musical genre and equalization settings.

The ideal candidate for the EQ problem is a metric that detects modifications in the equalization settings, but gives small distance values for similar songs equalized with the same settings.

General scheme of timbral similarity metrics



Tested timbral similarity metrics



Firstly the STFT of the signal *s* is computed and three features are retrieved from it:

- ***Ps*** or spectrogram
- ***Sonogram***
- ***MFCCs***

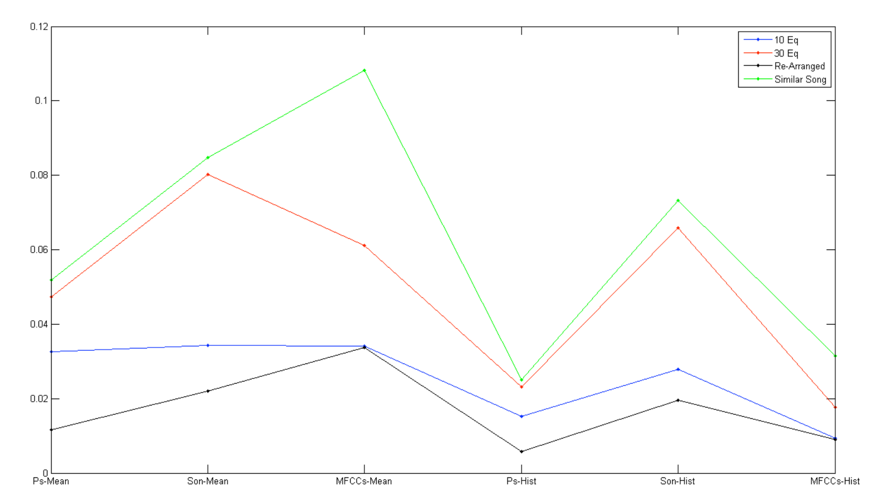
Then two different clustering techniques are used in order to find a feature that represent the entire song rather than a set of features for each window of the STFT.

- **Mean value**
- **Histogram**

This results in 6 different tested features.

Finally a normalized version of the Euclidean distance is used to assess the similarity between original and modified versions of the test song.

Future work



None of the tested metrics gives a smaller distance for the similar song than for the EQ30 song, which means that a GO based on these metrics is likely to converge to a solution where the tracks are strongly and wrongly equalized.

Further research will focus on other features and clustering techniques in order to find the right metric for the EQ problem.