Audio Engineering Society

Convention Paper

Presented at the 138th Convention
2015 May 7–10 Warsaw, Poland

Evaluation of an Algorithm for the Automatic Detection of Salient Frequencies in Individual tracks of Multi-track Musical Recordings

Christopher Dewey[1] and Jonathan P. Wakefield[2]

[1] University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
c.dewey@hud.ac.uk

[2] University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
j.p.wakefield@hud.ac.uk

ABSTRACT

This paper evaluates the performance of a salient frequency detection algorithm. The algorithm calculates each FFT bin maximum as the maximum value of that bin across an audio region and identifies the FFT bin maximum peaks with the highest five deemed to be the most salient frequencies. To determine the algorithm's efficacy test subjects were asked to identify the salient frequencies in eighteen audio tracks. These results were compared against the algorithm's results. The algorithm was successful with electric guitars but struggled with other instruments and in detecting secondary salient frequencies. In a second experiment subjects equalised the same audio tracks using the detected peaks as fixed centre frequencies. Subjects were more satisfied than expected when using these frequencies.

1. INTRODUCTION

The automatic detection of salient frequencies is of particular interest in the field of automated mixing and the development of tools designed to assist the engineer in performing a range of mix related tasks. Bitzer and LeBoeuf [1] have proposed one such solution which is based on the Welch periodogram. They attest that such a solution enables the engineer to spend more time on the creative aspect of mixing musical recordings. Cartwright et al [2] also seek to assist the engineer in equalization tasks and propose a solution that uses subjective audio descriptors to control EQ parameters. Gonzalez et al [3] employ a different approach and use cross-adaptive methods to perform automatic EQ in multiple audio tracks.

Previously the authors have proposed a number of visually intuitive novel equalization (EQ) tools [4]. These tools automatically detected and displayed the top five FFT bin maximum peaks for an audio region under consideration in order to assist the engineer in equalization tasks. These interfaces were evaluated by asking test subjects to remove a resonant frequency from an audio file consisting of a snare drum with an added resonant frequency. During this investigation, it was noted that none of the subjects questioned the frequencies presented via the interface. Whilst the algorithm did appear to be successful in detecting appropriate peaks for the EQ tasks during these, and other interface evaluation experiments, the previous paper was primarily focused on evaluating the novel equalization interfaces and didn't explicitly assess the performance of the FFT bin maximum peak algorithm. The aim of this paper is to present the algorithm and fully explore the algorithm's efficacy.

2. THE FFT BIN MAXIMUM PEAK DETECTION ALGORITHM

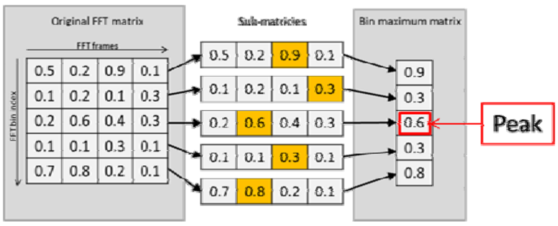

Figure 1 Identifying the bin maximums

The FFT bin maximum peak detection algorithm consists of two steps. Firstly each FFT bin maximum is calculated as the maximum value of that bin for all FFTs performed across the audio region under consideration as shown in Figure 1. The implementation uses a 1024-point FFT with a Hanning window function with 50% overlap. The second step is to identify the five FFT bin maximum peaks. A bin was considered a peak if its value was higher than the FFT bin either side of it for the audio region. The peaks were then sorted into magnitude order and the highest five were deemed the most salient frequencies. Clearly this is the simplest possible peak detection algorithm and a more sophisticated peak detection could be explored in future research.

A total of five peak frequencies were selected because traditional equalization interfaces on mixing desks typically contain four, five or six bands for EQ.

3. TEST AND EVALUATION

Two separate experiments were conducted to examine the efficacy of the FFT bin maximum peak detection algorithm for detecting salient frequencies. The first experiment was similar to LeBeouf and Bitzer's [6] and involved a group of subjects who were asked to identify salient frequencies in a broad range of tracks commonly found in popular musical recordings. These human identified frequencies could then be compared to the frequencies identified by the algorithm. In the second experiment subjects were asked to EQ the tracks presented in the first experiment using the five peak frequencies identified by the algorithm and to provide a user satisfaction score for how effectively they could equalize using just these frequencies.

3.1. Experiment 1

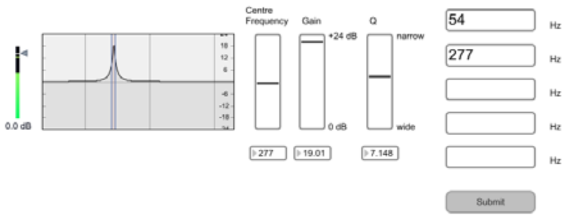

Figure 2 Interface used by subjects to identify salient frequencies

Twenty test subjects were asked to identify salient frequencies in eighteen tracks consisting of a selection of common instruments used in popular music. These consisted of bass guitar (2), kick drum (2), electric guitar (4), acoustic guitar (1) snare drum (2), male vocal (2), female vocal (1) and keyboards (4). Each track was been 3 and 10 seconds long and edited to loop continuously. Some tracks were well recorded and some tracks had obvious imperfections but were all considered 'usable' recordings by an independent expert mixing engineer subject to varying degrees of EQ processing. All test subjects were either second or third year undergraduate students on Music Technology or Music Production courses and deemed suitably experienced in mixing music to take part in this study.

A software interface was created in Max MSP for the experiment which provided the subjects with one band of parametric EQ to assist them in identifying up to five salient frequencies (see Figure 2). The order in which the audio samples were presented to the subjects was randomized. During the experiment the subjects were asked to use their own headphones. This meant that each subject could use means of monitoring the audio that they were familiar with. This approach also had the advantage of allowing a large number of subjects to could be tested in one session.

3.1.1. Results of Experiment 1

The results were compiled and compared against the output of the FFT bin maximum peak detection algorithm.

To aid interpretation a one dimensional plot was created for each track considered with the frequencies detected by the algorithm drawn as diamonds to the left of the plot and the subject's responses as coloured crosshairs as shown in Figure 3. These plots significantly aid interpretation and indicate that whilst the subjects appear to agree with some of the frequencies identified by the algorithm, there is a wide range of frequencies in each track that are deemed salient by the subjects yet remain unidentified by the algorithm. It is also noticeable that there appears to be a wide range of disparity between subjects with regard to the salient frequencies which was previously observed by Bitzer and LeBoeuf [1]. Furthermore, where there appears to be some consensus, the test subjects do not agree on the precise frequency of these saliencies.

As the FFT bin maximum peak detection algorithm uses a 1024-point FFT the identified peak frequency could in actuality lie anywhere between ±21.5 Hz from the bin's centre frequency i.e. a peak identified by the algorithm at 86Hz could be anywhere between 64.5 and 107.5Hz. Given this resolution the results were interpreted further with any subject identified frequency (SIF) found within the threshold of an algorithm identified frequency (AIF) deemed a positive match.

Figure 4 presents a histogram displaying the percentage of subjects that successfully matched each AIF in each track. Even in this analysis, there isn't strong agreement by the test subjects with the algorithm's detected frequencies. The algorithm appears to perform best with the bass guitar, muted electric guitar and distorted electric guitar tracks with the most dominant AIF being agreed upon by over 50% of subjects. In most cases the best match is with the 1st AIF.

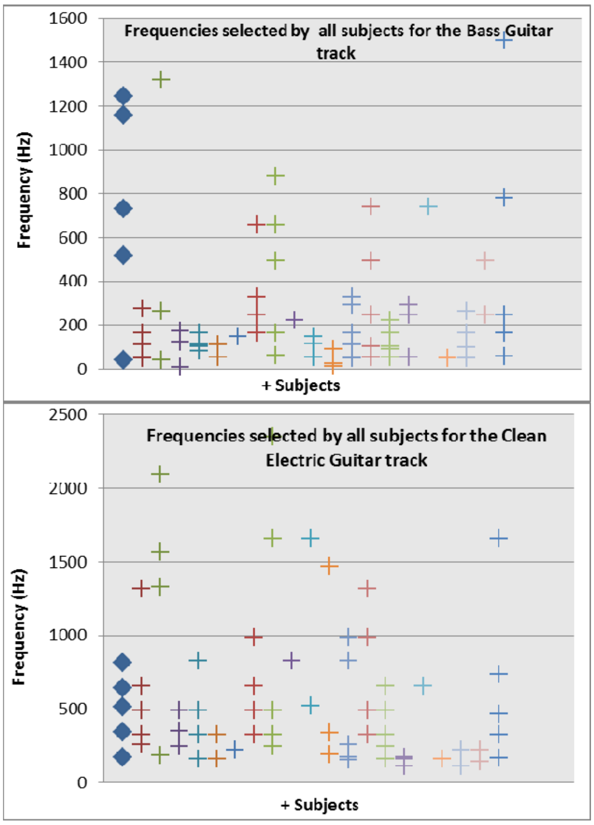

Figure 3 Example one dimensional plots showing the algorithm's detected frequencies and all subject identified frequencies for the Bass Guitar and Clean Electric Guitar tracks

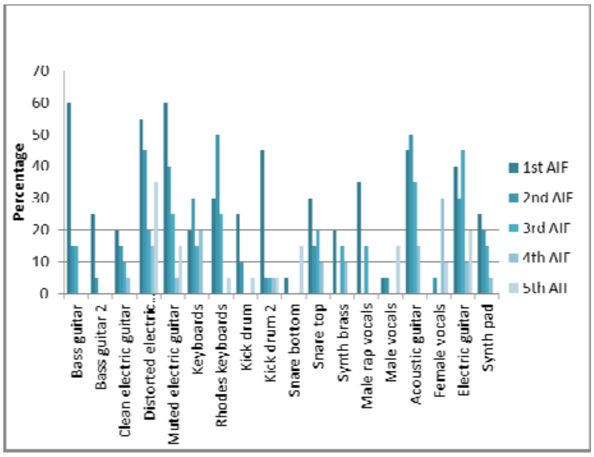

Figure 4 Percentage of subjects that identified each AIF in each track considered

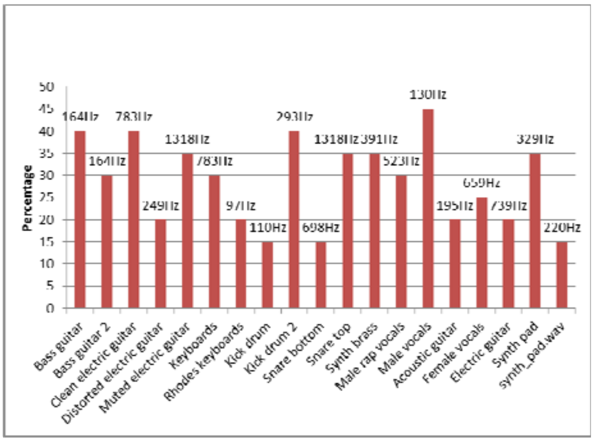

Figure 5 Percentage of subjects who agree on the most commonly identified non-AIF in each track

Figure 5 displays the most commonly identified non-AIF for each track as a histogram i.e. this shows us frequencies that subjects agreed upon that the algorithm did not identify. Clearly, in each track the algorithm fails to identify certain frequencies deemed salient by the subjects, however the highest percentage agreement across all tracks was 45%. Again this appears to show a lack of agreement between subjects.

Given that there is wide range of disparity between subjects with regard to the salient frequencies and where there appears to be some consensus the test subjects do not agree on the precise frequency of these saliencies this presents a difficulty in evaluating an algorithm. Consequently we decided to evaluate the efficacy of the algorithm's detected frequencies in a different way.

3.2. Experiment 2

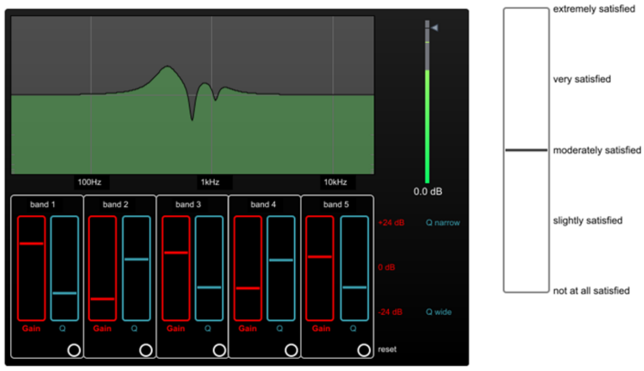

Figure 6 The EQ interface used in experiment 2

One week after the first experiment ten of the test subjects took part in the second experiment. The subjects were presented with a conventional five-band DAW style EQ and asked to EQ the same audio samples used in experiment one to their satisfaction. The interface was developed in Max MSP and enabled the subjects to modify each band's attenuation/gain and Q with the centre frequency of each band of EQ fixed to one of the top five AIFs. Following established convention the controls for each EQ band were ordered from low to high frequency from left to right and positioned below a graphical visualisation of the EQ curve to provide the subjects with visual feedback for the controls, as shown in Figure 6. The test subjects were asked to record their satisfaction with using this tool for each track using a sliding five-point Likert scale to the right of the EQ interface. A slider value of 1 corresponds to 'not at all satisfied' with a value of 5 corresponding to 'extremely satisfied'.

3.2.1. Results of Experiment 2

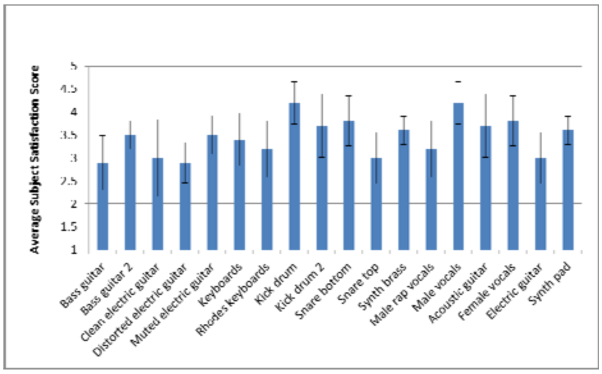

Figure 7 Histogram showing the average subject satisfaction score for all tracks

The results indicate that subjects were more satisfied using an EQ tool restricted to the AIFs to EQ the tracks than would be anticipated from the results of Experiment 1. Test subjects were on average 'moderately satisfied' (i.e. a score of 3) with the controls in 89% of the tracks under consideration (as shown in Figure 7). The subjects were on average 'very satisfied' using the AIFs to EQ the kick drum and male vocal tracks and malevox.wav tracks. This is of particular interest because both of these tracks appeared to score poorly in experiment 1. It is possible that by setting a wide Q, subjects were able to achieve satisfactory results with the controls provided despite pre-set centre frequency.

At the end of the experiment many of the subjects commented that the AIFs provided a "good starting point" from which to fine tune each EQ band's centre frequency. Interestingly, one subject commented that "if these were the only tools available to do the job you would get used to them."

4. CONCLUSIONS

A simple frequency detection algorithm which identifies the top five FFT bin maximum peaks of a single track is evaluated in terms of its efficacy in detecting salient frequencies presented.

Two experiments were conducted. The first involved comparing the frequencies identified by the algorithm against the frequencies identified by a group of test subjects for a range of audio tracks. In a second experiment test subjects were asked to use the frequencies identified by the algorithm within an EQ tool to EQ the same audio tracks and the subjects' satisfaction with the tool was recorded.

The results indicate that the algorithm appears to perform well in the identification of salient frequencies in recordings of some guitars when compared against the frequencies identified by test subjects but struggles with other recordings.

There is wide range of disparity between test subjects with regard to the salient frequencies and where there appears to be some consensus the test subjects do not agree on the precise frequency of these saliencies which presents a difficulty in evaluating an algorithm.

Subjects appear more satisfied on average than expected when using the frequencies identified by the algorithm in an EQ tool to perform EQ on a range of tracks.

5. REFERENCES

[1] Bitzer, J., & LeBoeuf, J. (2009, May). Automatic detection of salient frequencies. In Audio Engineering Society Convention 126. Audio Engineering Society.

[2] Cartwright, M. B., & Pardo, B. (2013). Social-EQ: Crowdsourcing an Equalization Descriptor Map. In ISMIR (pp. 395-400).

[3] Perez-Gonzalez, E., & Reiss, J. (2009, October). Automatic equalization of multichannel audio using cross-adaptive methods. In Audio Engineering Society Convention 127. Audio Engineering Society.

[4] Dewey, C., & Wakefield, J. (2013, May). Novel designs for the parametric peaking EQ user interface for single channel corrective EQ tasks. In Audio Engineering Society Convention 134. Audio Engineering Society.

[5] Bitzer, J., LeBoeuf, J., & Simmer, U. (2008, May). Evaluating perception of salient frequencies: Do mixing engineers hear the same thing?. In Audio Engineering Society Convention 124. Audio Engineering Society.