Towards Quantifying the “Album Effect” in Artist Identification

Youngmoo E. Kim, Donald S. Williamson, and Sridhar Pilli
Drexel University
Electrical and Computer Engineering
{ykim,dsw36,sp375}@drexel.edu

Abstract

Recent systems for automatically identifying the performing artist from the acoustic signal of music have demonstrated reasonably high accuracy when discriminating between hundreds of known artists. A well-documented issue, however, is that the performance of these systems degrades when music from different albums is used for training and evaluation. Conversely, accuracy improves when systems are trained and evaluated using music from the same album. This performance characteristic has been labeled the “album effect”. The unfortunate corollary to this result is that the classification results of these systems are based not entirely on the music itself, but on other audio features common to the album that may be unrelated to the underlying music. We hypothesize that one of the primary reasons for this phenomenon is the production process of commercial recordings, specifically, post-production. Understanding the primary aspects of post-production, we can attempt to model its effect on the acoustic features used for classification. By quantifying and accounting for this transformation, we hope to improve future systems for automatic artist identification.

Keywords: Artist identification, song classification, album effect, music production

1. Introduction

The explosion of digital music has led to unprecedented access to large and varied music collections, and now we are faced with the difficulty of organizing, labeling, and searching these libraries. This problem is exacerbated by the unreliability of music metadata, which is primarily contributed by the community at large and is susceptible to input errors and the preferences of individual users (due to the lack of globally adopted labeling standards). A large amount of research in music information retrieval has been focused on tasks which attempt to automatically extract relevant information about the music from the acoustic signal itself. Artist identification is one such problem that has been the subject of myriad papers as well as a competitive task in several of the recent MIREX events.

In early work in this area, it became apparent that artist classification scores using supervised learning systems improved when songs from the same album were used for both training and evaluation [1]. Conversely, it has been shown that the performance of most systems degrades substantially when the albums used in training and testing are mutually exclusive [2]. This performance characteristic was first coined by Whitman *et al* as the “album effect”.

The majority of artist identification systems employ acoustic features modeling the spectral characteristics of audio (such as Mel-frequency cepstral coefficients and linear prediction coefficients). It follows that the album effect is primarily the result of frequency-domain features common to the songs on a given album. Although there are notable exceptions, albums tend to be recorded within a relatively short consecutive time period, so it makes sense that an artist’s choice of instrumentation is fairly consistent within an album while it varies somewhat more between albums. Perhaps even more important, however, are the aesthetic sensibilities of the producer(s), who has overall creative control of the album. The producer’s choices in equipment, orchestration, and particularly audio effects and post-production will impart a spectral imprint on the overall album. Hence, the album effect has since come to also be known as the “producer effect”.

2. Commercial music production

Production of an album consists of multiple stages, each of which contributes to the overall album effect. The choice of recording studio and equipment will play a significant role since they will be consistent across all songs of an album (and will likely change between albums of a given artist). During recording, microphones will impart characteristic frequency responses that will affect the spectrum of the recorded sound. Studio consoles employ different electronics and filters, which will also impart their signatures to the sound. And during mixing the choice of sound monitors, used by the producer as a reference for the output, will also have a direct effect on the final recording.

It is likely that the producer’s aesthetic choices during mixing also contribute to the album effect. For example, the producer may employ a consistent *equalization*, filtering to boost or attenuate particular frequency ranges, to the lead vocalist as well as other instruments. Originally implemented using analog filters, equalization (or EQ) is now systematically applied to most sound elements in a song using digital filters. Across the album there is usually consistency in the placement of instruments in a stereo field (panning). Additionally, the wideness (or lack of wideness) of the stereo image is generally the same for all of the songs. A producer's continued use of the same digital audio effects, such as reverberation, chorus, flanging, etc., will also provide the songs with more commonalities.

Post-production consists of processing that is applied to all of the individual songs of an album after they have been mixed to provide a consistent sound quality across the album. This can include the application of additional equalization, dynamics, and effects. *Dynamics* refer to dynamic compression or expansion, i.e. manipulation of the dynamic range of the audio. This is done during post-production to ensure that songs on an album do not dramatically change in volume so that one can listen to the entire album without changing the level. In popular music, dynamic compression is usually applied so that playback on lower-quality audio speakers will not result in distortion.

3. Acoustic Features used in Classification

The most common acoustic features used in automatic classification systems are mel-frequency cepstal coefficients. These employ the well-known *cepstrum* (inverse-log-Fourier transform), but weighted according to the mel-frequency scale to approximate the frequency resolution of human perception. A relatively small number of MFCCs approximate the spectral envelope of a short-time segment of a signal.

MFCCs are ordered in terms of the bandwidth of energy they contribute to the signal, from wideband to narrowband. Thus, the lowest coefficient is correlated with the full-band energy contained within the signal (approximating the amplitude envelope). Higher-order coefficients add more detail to the spectral envelope approximation. The representation can be inverse transformed to a spectral representation, so it is easy to gauge the relative contributions of individual coefficients.

Equalization is the most obvious post-production feature reflected in the MFCCs. A post-production equalization filter applied to all songs within an album will bias the MFCCs in a consistent manner. The dynamic range of the music (indicating the amount of dynamic compression applied) will also be indicated by the first mel-cepstrum coefficient.

4. Analysis Examples

A particularly interesting data set is formed from original recordings of hit songs and re-releases of those songs as part of "Greatest Hits" albums. In this specific case, the exact recording of the original hit is almost always used for the re-release (the intent, after all, is to capitalize on the popularity of the earlier release). These collections are usually *remastered*, meaning that the group of hit songs (from multiple albums) undergoes additional post-production to create a consistent sound quality across the new collection. With older songs, the recordings may also be re-digitized from analog tape with more modern digital-to-analog converters with improved characteristics. Using this data set, we can compare the characteristics of two different masterings of the same recording.

If we examine the MFCCs across an entire song and average the resulting time-varying spectral envelope we observe distinct differences between an original recording and its remastering (Figure 4), which appears to be the result of equalization.

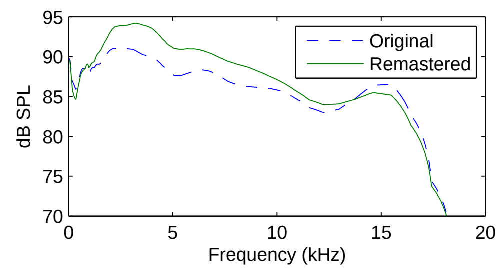

Figure 1. Average MFCC spectra of original and remastered recordings of *The Unforgettable Fire* by U2.

The distribution of power per analysis frame for each version of the song reveals that the remastered version has a higher overall dynamic level, as well as a wider dynamic range indicating another post-production alteration.

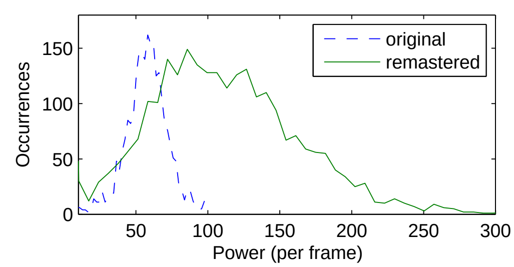

Figure 2. Distribution of signal power per frames of original and remastered recordings of *The Unforgettable Fire* by U2.

5. Future Directions

We intend to use these results to model the within-class variance for each artist, in order to identify features that are more indicative of the production as opposed to the music of the artist. In this way, we hope to improve the overall performance of artist classification systems.

References

[1] B. Whitman, G. Flake, and S. Lawrence. "Artist detection in music with Minnowmatch," in *Proc. IEEE Workshop on Neural Networks for Signal Processing*, 2001, pp. 559-568.

[2] M.I. Mandel and D.P.W. Ellis. "Song-level features and support vector machines for music classification," in *Proc. Int'l Sym. on Music Information Retrieval 2005*, pp. 594-599.