Segmenting electronic dance music streams based on self-similarity, symmetry, evolution and contiguity

Tim Scarfe, Wouter M. Koolen and Yuri Kalnishkan
Computer Learning Research Centre and Department of Computer Science,
Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom
{tim,wouter,yura}@cs.rhul.ac.uk

April 1, 2014

Abstract

We describe an unsupervised, deterministic algorithm for segmenting DJ-mixed Electronic Dance Music (EDM) streams (for example; pod-casts, radio shows, live events) into their respective tracks. We attempt to reconstruct a fixed number of boundaries as close as possible to what a human domain expert would assign in respect of the same task. The goal of DJ-mixing is to render track boundaries effectively invisible from the standpoint of human perception which makes the problem difficult.

We use optimally segment domain inspired cost matrices derived from a self-similarity matrix. Dynamic programming is used (which means solutions to a problem are described in terms of overlapping sub-solutions to achieve a significant improvement in time complexity and therefore execution time). The similarity matrix is based on the normalized cosines of a time series of transformed Fourier based features. Our method is applied to dance music radio shows. Its formulation incorporates long-term self similarity over a variable (but limited) horizon as a first class concept combined and it is quantitatively assessed on a large corpus of 304, two hour long radio shows which have been hand-labelled by a domain expert.

Keywords. music, segmentation, DJ, mix, dynamic programming

1 Introduction

Electronic Dance Music tracks are usually mixed by a disc jockey (DJ). For this reason EDM music streams are unique compared to other genres of music. Mixing is the *modus operandi* in electronic music. We first transform the audio file into a time series of features discretized into adjacent tiles and transform them into a domain where some pairs from the same track would be distinguishable by their cosine. Our features are based on a Fourier transformation with convolution filtering to accentuate prominent instruments and self-similarity within tracks. We create a similarity matrix from these cosines and then derive cost matrices showing the costs of fitting a track at a given time with a given length. We use Dynamic Programming to create the cost matrices and again to perform the most economical segmentation of the cost matrices to fit a predetermined number of tracks.

A distinguishing feature of our algorithm is that it focuses on long term self similarity of segments rather than transients or a novelty function. Dance music tracks have the property that they are made up of repeating regions, and the ends are almost always similar to the beginning. For this reason we believe that some techniques from structural analysis fail to perform as well for this segmentation task because we focus on the concept of self-similarity ranging over a configurable time horizon. Our method does not require any training or tenuous heuristics to perform

well.

The purpose of this algorithm is to reconstruct optimal boundaries given a fixed number of tracks known in advance (their names and order are known). This is relevant when one has recorded a show, downloaded a track list and needs to reconstruct the indices given a track list. The order of the indices reconstructed is critical so that we can align the correct track names with the reconstructed indices. If the track list were not known in advance the number of tracks could be estimated quite reliably as the variable of track lengths is Gaussian (see Figure 1).

One of the interesting features of audio is that you *cannot scrub through it, and get an overview in the same way you can with video.* Audio has a reduced *contextual continuum* when the user skips through it, perhaps due to the lack of redundant, persistent scene-setting information or indeed a psychological reason. Even in video applications, discovery, context and scrubbing are an active area of research [1]. Time index metadata would allow click through monetisation, and allow improved use-case scenarios (for example publishing track names to social networks, information discovery and retrieval). Capturing metadata in audio is a time consuming and error-prone process. Tzanetakis [2] found that it took users on average 2 hours to segment 10 minutes of audio using standard tools. While not directly relevant we might glean from those findings that there is a strong motivation to automate this process. We have also noticed that while the error variable of the human captured time metadata that we were supplied with appears Gaussian, there is a noticeable deterioration towards the end of the shows or in relation to the complexity of the mixing by the disc jockey.

To mix tracks DJs always match the speed or *BPM* (beats per minute) of each adjacent track during a transition and align the major percussive elements in the time domain. This is the central concept of removing any dissonance from overlapping tracks. Tracks can overlap by any amount. DJs increase adjacent track compatibility further by selecting adjacent pairs that are harmonically compatible and by applying spectral transformations (EQ).

The main theme of the early literature was attempting to generate a novelty function to find points of change using distance-based metrics or statistical methods. Heuristic methods with hard decision boundaries were used to find the best peaks. A distinguishing feature of our approach is that we evaluate how well we are doing compared to humans for the same task. We compare our reconstructed indices to the ones created by a human domain expert.

J. Foote et al [3, 4, 5, 6, 7] have done a significant amount of work in this area and were the first to use similarity matrices. Foote evaluated a Gaussian tapered checkerboard kernel along the diagonal of a similarity matrix to create a 1d novelty function. One benefit to our approach is that our DP allows any range of long-term self similarity (which relates to the fixed kernel size in Foote's work).

Goodwin et al. also used DP for segmentation [8, 9]. Their intriguing supervised approach was to perform Linear Discriminant Analysis (LDA) on the features to transform them into a domain where segmentation boundaries would be emphasised and the feature weights normalized. They then reformulated the problem into a clustering DP to find an arbitrary number of clusters. We believe the frame of mind for this work was structural analysis, because it focuses on short term transients (mitigated slightly by the LDA) and would find segments between two regions of long term self similarity. Goodwin was the first to discuss the shortcomings of novelty peak finding approaches. Goodwin's approach is not optimized to work for a predetermined number of segments and depends on the parametrization and training of the LDA transform.

Peeters et al [10, 11] did some interesting work combining k-means and a transformation of the segmentation problem into Viterbi (a dynamic program).

We compare our error to the relative error of cue sheets created by human domain experts. We focus directly on DJ mixed electronic dance music.

In the coming sections we describe the corpus (see Section 2), the evaluation criteria (see Section 3), how we pre-process the data (see Section 4), how we perform feature extraction (see Section 5), designing the cost matrices using observed phenomena in the domain (see Section 6), computing the best segmentation (see Section 7), discussion of confidence intervals

(see Section 8), our methodology (see Section 9), materials used (see Section 10), results (see Section 11) and finally the summary (see Section 12).

2 Corpus

We have been supplied with several broadcasts from three popular radio shows. These are: Magic Island, by Roger Shah (106 shows); A State of Trance with Armin Van Buuren (110 shows); and Trance Around The World with Above and Beyond (88 shows) (Total 304 shows). The show genres are a mix of Progressive Trance, Uplifting Trance and Tech-Trance. We believe this corpus is the largest of its kind used [12] in the literature. The music remains uninterrupted after the introduction (no silent gaps). The shows come in 44100 samples per second, 16 bit stereo MP3 files sampled at 192Kbs. We resampled these to 4000Hz 16 bit mono (left+right channel) WAV files to allow us to process them faster. We have used the "Sound eXchange"[1] program to do this. These shows are all 2 hours long. The overall average track length is 5 and a half minutes (slightly less for Magic Island (see Figure 1)) and normally distributed. The average number of tracks is 23 for ASOT and TATW, 19 for Magic Island. There is a guest mix on the second half of each show. The guest mix DJs show off their skills with technically convoluted mixing.

There is already a large community of people interested in getting track metadata for DJ sets. "CueNation"[2] is an example of this. CueNation is a website allowing people to submit *cue sheets* for popular DJ Mixes and radio shows. A cue sheet is a text file containing time metadata (indices) for a media file.

We had our time metadata (indices) and radio shows provided to us and hand captured by *Dennis Goncharov*; a domain expert and one of the principal contributors to *CueNation*. One of the significant problems with this task is that there is an (apparently Gaussian) error variable attached to the human captured indices themselves. On some tracks, it is unclear where to place the optimal index and when analysing our results, we have noticed some obvious

[1] `http://sox.sourceforge.net`
[2] `http://www.cuenation.com`

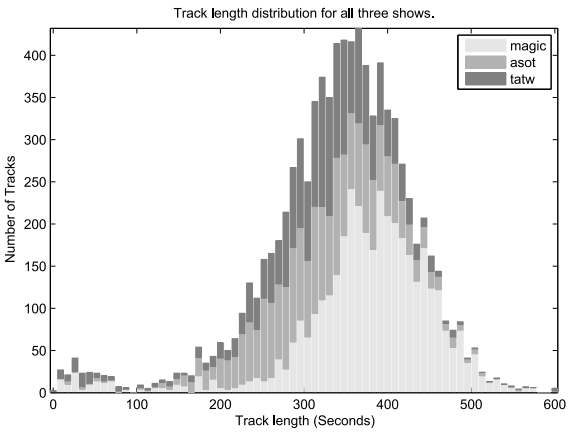


Figure 1: The track length distribution for all three radio shows. The *bump* of short tracks (less than 3 minutes) is often introductions or otherwise extraneous.

human errors. Regrettably, there is no clear way of quantifying this. Many of the cue sheet authors themselves reject the idea of automating the task citing the poor precision of any such result (they often place indices on the exact MP3 frame). However this sentiment seems misplaced given that they make blatant mistakes or that it is a matter of opinion where to place the track and some consistent method may be preferential. A potential outcome of our method could be a way to assist them with initial placements. Our results demonstrate that it is indeed possible to automate this task and that while there is some uncertainty attached to the optimal placement, it is still largely predictable.

Dennis Goncharov provided us with the following description of how he captures the indices. To quote from a personal email exchange with Dennis:

> The transition length is usually in factors of 8 bars (1 bar is 4 beats. At 135 beats per minute, 8 bars is 14.2 sec). It is a matter of personal preference which point of the transition to call the index. My preference is to consider the index to be the point at which the second track becomes the focus of attention and the first track is sent to

the background. Most of the time the index is the point at which the bass line (400Hz and lower) of the previous track is cut and the bass line of the second track is introduced. If the DJ decides to exchange the adjacent tracks gradually over the time instead of mixing them abruptly then it is up to the cue sheet maker to listen further into the second track noting the musical qualities of both tracks and then go back and choose at which point the second track actually becomes the focus of attention.

The most obvious and pervasive element in dance music is the percussion (the beats). We believe on balance that ignoring the percussive information is advantageous, because DJs use percussion primarily to blur boundaries between tracks. We tried to capture percussive based features and found that the transitions between tracks and indeed groups of tracks appeared as stronger self-similar regions in S than the actual tracks. It is clear from our research that it is the boundaries between instruments and harmonic content which reveal track boundaries, not percussion or rhythm. When we tried implementing a rhythm feature extractor by looking at a transformed autocorrelation of the time domain tiles, it was ineffective because it is the rhythm more than the instruments that is matched between adjacent tracks by the DJs. Some DJs mix harmonically (by matching instruments as opposed to percussion) but this preys on human hearing and perception. An algorithm capturing the harmonic information would most likely be able to distinguish two harmonically compatible tracks. Perhaps more sophisticated rhythm detection (than we tried) might still work and should be explored further.

3 Evaluation Criteria

It is challenging to quantify the performance of our method because if we misplace any tracks, it may have a cascade effect. For example if we place one track too many early on in a show, many of the subsequent tracks may be correctly detected but placed out of alignment.

We perform three types of evaluation: '*average*', *heuristic precision*', and '*thresholds*'.

The *average* (in seconds) given as

$$\hat{A}(P, A) = \frac{1}{|P|} \sum_{i=1}^{|P|} |P_i - A_i|$$

is quite simply the mean absolute difference between constructed and actual indices (P is constructed indices, and A is the human indices). Naturally this metric will have a reduced meaning in some sense because although it depends on the real accuracy it also depends on the number of misplacements and where in the track those misplacements were. The uncertainty variable attached to the misplacements is wide (in relation to the size of the shows and the number of tracks) but does regularise over the course of a large dataset. This metric is the most meaningful for capturing the overall *robustness* of the method.

The *heuristic precision* metric denoted by

$$\hat{H}(P, A) = \text{Median}\left(\text{Sort}\left(|P_i - A_{1,2,\ldots,|P|}|\right)_1\right)$$

for all $i = 1, 2, \ldots, |A|$ takes the median of the absolute differences between each prediction in P and the nearest actual index in A. This is the most meaningful single metric of accuracy. It is largely invariant to misplacements.

The thresholds metric is the percentage of matched tracks within different intervals of time

$$\{\ 60,\ 30,\ 20,\ 10,\ 5,\ 3,\ 1\ \},$$

in *seconds* as a margin around any of the track indices we have been given. This metric is also invariant to alignment and is provided for comparison with our other paper [13].

Note that our results will contain the mean average of these metrics (captured for each show) across the dataset being evaluated.

4 Preprocessing

The dataset had some outliers that may have slightly distorted the analysis of our method. Many of the "tracks" in our data set (of indices) were in fact not

tracks at all but rather introductions or voice-overs. Almost all of these outlier tracks were short in length. These are quite clearly visible on the distribution of track lengths on Figure 1. To ameliorate the situation we removed any tracks that were shorter than 180 seconds. We also removed any end tracks that were shorter than 240 seconds as very often the end tracks on a radio show contain strange elements (for example voice-overs, interviews, show-related 'jingles'). This required some manipulation of the cue sheets and audio files. The undesired segments of the audio files were chopped out, and the cue sheets were reflowed so that the time indices point to the correct location in the file.

The algorithm still performs similarly when removing just these indices and leaving the audio intact underneath, so it would not significantly affect any real-world implementation.

For those wishing to use this algorithm in practice with pre-recorded shows; the introductions at the start of the shows can be thought of as being fixed length (with a different length for each show type).

5 Feature Extraction

We used SoX (see Sect. 2) to downsample the shows to 4000Hz. We are not particularly interested in frequencies above around 2000Hz because instrument harmonics become less visible in the spectrum as the frequency increases. The Nyquist theorem [14] states that the highest representable frequency is half the sampling rate, so this explains our reason to use 4000Hz. We will refer to the sample rate as R. Let L be the length of the show in samples.

Fourier analysis allows one to represent a time domain process as a set of integer oscillations of trigonometric functions. We transform the tiles into the frequency domain using the discrete Fourier transform

$$F(x_k) = X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi \frac{k}{N} n}$$

which transforms a sequence of complex numbers $x_0, \ldots, x_N$ into another sequence of complex numbers $X_0, \ldots, X_N$ where

$$e^{-i2\pi \frac{k}{N} n}$$

are points on the complex unit circle. Note that the fftw algorithm [15] that we used to perform this computation operates significantly faster when N is a power of 2 so we zero pad the input to the next power of 2. We denote the tile width by M in seconds (an algorithm parameter). Note that

$$N = \frac{L}{M}$$

denotes the tile size in samples (length of show in samples over the tile size). Let

$$T = \left\lfloor \frac{L}{\tilde{M}} \right\rfloor$$

be the total number of tiles, and

$$\tilde{M} = \frac{L}{N}$$

the tile width in samples. Because we are passing real values into the $F(x_k)$, the second half of the result is a rotational copy of the first half.

Show samples are collated into a time series Q_i^y $(T \times N)$ of contiguous, non-overlapping, adjacent *tiles* of equal size where $i = 1, 2, \ldots, T$. Samples at the end of the show that do not fill a complete tile get discarded. The affect of this is increasingly negligible with decreasing tile size. Since we zero-pad N to the next power of two, this also decreases the affect.

As we are not always interested in the entire range of the spectrum, we use l to represent a low pass filter (in Hz) and h the high pass filter (in Hz). So we will capture the range from h to l on the first half of the result of F. Let $\hat{h} = \lceil\ h\ \cdot\ \frac{N}{R}\ \rceil + 1$ be the position of h in the spectrum, and $\hat{l} = \lceil\ l\ \cdot\ \frac{N}{R}\ \rceil + 1$ be the position of l in the spectrum.

Let D_e^y $(T \times \hat{l} - \hat{h} + 1)$ denote the feature matrix. For each tile $i* = 1, 2, \ldots, T$ we assign

$$D_{i*}^{1,\ldots,\hat{l}-\hat{h}+1} = \left|\ F(Q_{i*}^{1,\ldots,\tilde{M}})_{\hat{h},\ \hat{h}+1,\ \ldots,\ \hat{l}}\ \right|$$

selecting the part of the spectrum between the high and low pass filters h and f. We take the absolute

values of the complex result of $F(x_k)$ (defined as its distance in the complex plane from the origin using the Pythagorean theorem).

To focus on the instruments and improve performance we perform convolution filtering on the feature vectors in D, using a Gaussian first derivative filter. This works like an edge detection filter but also expands the width of the transients (instrument harmonics) to ensure that feature vectors from the same song appear similar because their harmonics are aligned on any distance measure (we use the cosines). This is an issue because of the extremely high frequency resolution we achieve from having such large inputs into $F(t_i)$. For example with a tile size of 10 seconds and a sample rate of 4000 we have a frequency resolution of $\frac{1}{2}10 \cdot 4000 = 20\text{KHz}$.

Typically a 'short-time discrete Fourier transform' is used which has smaller sized inputs (windows) into $F(t_i)$ which are usually overlapping and are multiplied by a window function, attenuating the tails to reduce spectral leakage. Usually these window functions look similar to a Gaussian, for example;

$$\text{Hann}_i = 0.5 - 0.5\cos\frac{2\pi i}{n-1}w(i)$$

where n is the window size (see [16] for an example). The short-time Fourier transform is relevant when increased time precision is needed as there is a frequency-time resolution trade-off with respect of the input size to $F(t_i)$. This is not a concern in this particular application as our time resolution is never required to be better than 1 second which would still give us adequate frequency resolution.

The Gaussian first derivative filter is defined as

$$-\frac{2\hat{\lambda}}{\upsilon^2}e^{-\frac{\hat{\lambda}^2}{\upsilon^2}}$$

where

$$\hat{\lambda} = \{\ -\lfloor 2\upsilon \rfloor, \lfloor -2\upsilon + 1 \rfloor, \ldots, \lfloor 2\upsilon \rfloor\ \},$$

and

$$\upsilon = b\frac{N}{R}.$$

b is the bandwidth of the filter in Hz and this is a parameter of the algorithm. After the convolution filter is applied to each feature vector in D, we take the absolute values and normalize on the vector lengths.

Because the application domain is well defined in this setting, we can design features that look specifically for what we are interested in (musical instruments). Typically in the literature; algorithms use an amalgam of general purpose feature extractors. For example; spectral centroid, spectral moments, pitch, harmonicity [2]. We construct a dissimilarity matrix of cosines as is common in the literature for similar applications [3]. The cosines are quickly computable because they are the the inner products of the respective features (the features have been normalized to unit length).

Let

$$S\ (T \times T)\ =\ 1 - \langle\ D_i \cdot D_j\ \rangle,$$

define the dissimilarity matrix.

We center S around its mean

$$S = S^{2 \cdot \frac{1}{T^2} \sum_{i=1}^{T} \sum_{j=1}^{T} S_{ij}},$$

then normalize it by $\hat{c}$ (an algorithm parameter), $S = S^{\hat{c}}$ and place it on a small interval around 0. $S_{ij} = (2 \cdot S_{ij}) - 1$. It is necessary to have *incentives* and *disincentives* for meaningful track placement which is why we opted for this interval. The distribution of values in S is normal with a raised section on left (negatives) representing the tracks we want to find. When $\hat{c}$ is 1 there is no change, smaller values will shift the mean above 0 and vice versa. This will become relevant when we discuss cost matrices as some of them depend on the sign of the value in S.

See Figure 3 for an illustration of S and Figure 2 for an illustration of the cosine normalization.

6 Cost Matrices

We now have a similarity matrix $S(i, j)$ as described in Section 5.

Let w and W denote the minimum and maximum track length in seconds, these will be parameters.

We require a cost matrix $C(f, t)$ that describes the cost of placing a track of length $t - f + 1$ at f. After analysing the data set, we have created 7 cost

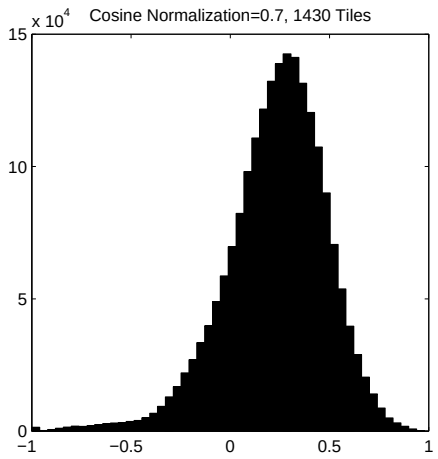


Figure 2: Illustration of the effect of normalization parameter $\hat{c} = 0.7$ on the values in S on radio show Magic Island 110. The small raised section on the left correspond to the tracks.

matrices that exploit observed phenomena in S. We also provide an additional cost matrix which is just a Gaussian random function centred around the mean track length for all times which can be used to regularise the other 7 matrices or used on its own as a comparator to a more naive method of placement.

The more sophisticated matrices exploit themes such as contiguity, symmetry, evolution and change as well as simple summation of S as was presented in our last paper. In our previous work [13] S was on the interval $[0, 1]$ and the summation method could only consider disincentives. All of the new matrices have a parameter to shift the consideration of incentive versus disincentive and will contain values on the interval $[-1, 1]$.

Intuitively, most tiles within the same track are similar, while pairs of tiles that do not belong to the same track are significantly more dissimilar. However, often this is not the case and many tracks have dissimilar regions within them.

6.1 Normalization

We define $N(C, \Omega)$ as a normalization function for cost matrix C and incentive bias Ω. It will return a $[-1, 1]$ interval when $\Omega = 0.5$, $[-1, 0]$ when $\Omega = 1$ and $[0, 1]$ when $\Omega = 0$ and will produce a convex interpolation for other values.

Let

$$\hat{N}(C, \Omega) = \left(\frac{C_{ij} - \min(C_{ij})}{\max(C_{ij}) - \min(C_{ij})} \cdot \hat{h} \right) - s$$

for all $i, j \in C$ where $\hat{h} = 2 - (2 \cdot |0.5 - \Omega|)$ and

$$s = \begin{cases} 1 - 2 \cdot |0.5 - \Omega| & \text{if } \Omega > 0.5 \\ 1 & \text{otherwise .} \end{cases}$$

Note that to save space we will use the following notation $\hat{N}(C, \Omega) = \hat{N}_{\Omega}(C)$.

6.2 Summation

The most obvious strategy of all is to sum up all relevant tiles in S for each candidate track from tile f through tile t. We define $C(f, t)$, the cost of a candidate track from tile f through tile t, to be the sum of the similarities between all pairs of tiles inside it, normalized on track length.

$$C(f, t, \Omega) = \frac{\sum_{i=f}^{t} \sum_{j=f}^{t} \hat{S}_{\Omega}(i, j)}{(t - f + 1)}$$

where

$$\hat{S}_{\Omega} = \begin{cases} \Omega S_{ij} & : S_{ij} > 1 \\ (1 - \Omega) S_{ij} & : \text{otherwise} \end{cases}$$

for all $i, j \in S$. Direct computation using the definition takes $O(TW^3)$ time. We can improve this to $O(TW^2)$ using the recursive formulation for the unnormalized quantity

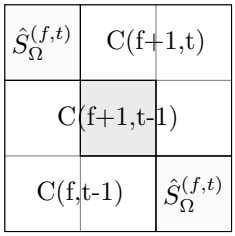
$$\begin{aligned} &\tilde{C}(t - (w + 1), t + (w - 1)) = \\ &\tilde{C}(t - (w - 1), t + (w - 1)) \\ &\quad + \textstyle\sum_{i=1}^{w} \tilde{C}(t - (w - 1), t + (w - 1) + i) \end{aligned}$$

for all $t = 1, \ldots, T$ and $w = 1, \ldots, \min(w, t)$.

$O(TW)$ is achievable using the following recursion for the unnormalized quantity $\tilde{C}(f,t) = C(f,t)(t-f)$ (for $f+1 \leq t-1$)

$$\begin{aligned}\tilde{C}(f,t) &= \tilde{C}(f+1,t) + \tilde{C}(f,t-1) \\ &\quad - \tilde{C}(f+1,t-1) + \hat{S}(f,t) + \hat{S}(t,f).\end{aligned}$$

(having first pre-computed C for each $1 \leq f \leq t \leq T$). This can be better understood with the following illustration



The final cost matrix is normalized by width and incentive bias

$$C = \hat{N}_{\Omega}\left(\frac{\tilde{C}_{ft}}{t-f+1}\right)$$

for all $t, f \in \tilde{C}$.

6.3 Symmetry

A common feature on dance music tracks is partial mirror-symmetry.

Let

$$\begin{aligned}\Lambda(f,t,d) = \{\ & S(f+d-1,f), S(f+d,f+1), \\ & S(f+d+1,f+2), \ldots, S(t,t-d-1)\ \}\end{aligned}$$

take the diagonal d from track

$$\hat{t} = S(f, f+1, \ldots, t, f, f+1, \ldots, t).$$

For each diagonal in one triangle/half of $\hat{t}$ (it is a symmetric matrix) we want to compare each element against its mirror counterpart. 'Symmetric' pairs that contain the same sign are multiplied together. Pairs that have a different sign are set to 0.

Let the symmetry cost matrix

$$C = \frac{\hat{N}_{\Omega}(\ \tilde{C}(f,t,\Omega,\delta)\ )}{(t-f+1)}$$

using the following definition for $\tilde{C}$

$$\tilde{C} = \sum_{d=1}^{t-f+1} \sum_{i=1}^{|\Lambda(f,t,d)|} \tilde{s}(\Lambda(f,t,d)_i, \Lambda(f,t,d)_{|\Lambda(f,t,d)|-i-1})$$

where

$$\tilde{s}(p,q) = \begin{cases} 0 & : \text{sign}(p) \neq \text{sign}(q) \\ \delta(p,q,\Omega) & : \text{otherwise.} \end{cases}$$

We will consider 3 types of symmetry cost matrix. The *standard symmetry* version where pair evaluator

$$\begin{aligned}&\delta(p,q,\Omega) \\ &= \begin{cases} (\Omega)(p \cdot q) & : \max(\text{sign}(p), \text{sign}(q)) = 1, \\ (1-\Omega)(p \cdot q) & : \quad \text{otherwise} \end{cases}\end{aligned}$$

symmetry summation where:

$$\begin{aligned}&\hat{\delta}(p,q,\Omega) \\ &= \begin{cases} (\Omega)\frac{1}{2}(p+q) & : \max(\text{sign}(p), \text{sign}(q)) = 1 \\ (1-\Omega)\frac{1}{2}(p+q) & : \text{otherwise} \end{cases}\end{aligned}$$

And finally the symmetry change matrix is defined where $\hat{\delta}$ is the same as the *symmetry summation*, the width normalization is removed and the inner sum of the definition becomes the first order differences. For the sake of brevity we have omitted the dynamic programming formulations which run in $O(TW^2)$ time but they have been implemented in code and are available on-line (see Section 10). This is also the case for all the following cost matrices.

6.4 Static Contiguity

Horizontal contiguous traces in S indicate that the track is self-similar (low value ≤ 0) or self-dissimilar (≥ 0) due to repetition with the additional information that it is not evolving with respect of time either. If a given tile is the same as a set of contiguous tiles following it, we can assume that there is some static contiguous region. We define contiguity parameter ρ to indicate how many contiguous tiles are required. Let

$$\Gamma(f,t,h) \quad = S(\ f+h-1,\ \tilde{i}(f,t,h)\ )$$

define the future self similarity trace for track between f and $f+t-1$ and from time horizon h where

$$\tilde{i}(\ f,t,h\ ) = \{\ f+h-1, f+h, \ldots, t-h-1\ \}.$$

We define the future contiguity quantity C

$$C(f,t) = \hat{N}_\Omega \left(\frac{\tilde{\sum}_{h=1}^{t-f+1} \sum_{i=\rho}^{|\Gamma(f,t,h)|} \tilde{x}\left(\Gamma(f,t,i)_{\tilde{h}}\right)}{t-f+1} \right)$$

where

$$\tilde{h} = \{i-\rho, i-\rho+1, \ldots, i\},$$

$$\tilde{x}(v) = \begin{cases} \tilde{\delta}(v,\Omega) & : \text{range}(\tilde{n}(v)) = 0 \\ 0 & : \text{ otherwise,} \end{cases}$$

$$\tilde{n}(v) = \{\ sign(v_1), sign(v_2), \ldots, sign(v_{|v|})\ \},$$

$$\tilde{\delta}(v,\Omega) = \frac{1}{|v|} \sum_{x=1}^{|v|} \tilde{f}(\Omega, v)$$

and

$$\tilde{f}(\Omega, v) = \begin{cases} \Omega \cdot v_x & : \text{ sign}(v_1) = 1 \\ (1-\Omega)\cdot v_x & : \text{ otherwise.} \end{cases}$$

We will also create a past contiguity cost matrix where we replace $\Gamma(f,t,h)$ with a version

$$\hat{\Gamma}(f,t,h) \quad = S(\ \hat{i}(f,t,h), f+h-1\ ),$$

where

$$\hat{i}(\ f,t,h\ ) = \{\ t-h-1, t-h, \ldots, f+h-1)\ \}.$$

which traces past self similarity from the end of the track being fitted.

6.5 Evolutionary Contiguity

Any diagonal traces in S that are parallel to the main diagonal are partial copies of the track in the future which evolve in respect of time. Evolutionary Contiguity is a cost matrix which compares all adjacent pairs on the diagonals in $\hat{t}$, using the comparator $\delta(p,q,\Omega)$ from the standard symmetry cost function and multiplies those values by the time horizon. All the values are summed and normalized by the track width squared.

Let the evolutionary cost matrix

$$C(f,t,\Omega) = \hat{N}_\Omega \left(\frac{\sum_{d=1}^{t-f+1} \sum_{i=2}^{|\Lambda(f,t,d)|} \delta(\kappa_i, \kappa_{i-1}, \Omega)\cdot d}{(t-f+1)^2} \right)$$

where $\kappa_i = \Lambda(f,t,i)$. As before $\tilde{C}$ is normalized by track width and incentive bias.

6.6 Gaussian

Let

$$G(\varpi, N)_{tw} = e^{-\frac{1}{2}\frac{\varpi n}{\frac{1}{2}W}^2}$$

for all $n = 1, 2, \ldots, W$ denote the Gaussian matrix cost function of $N \times W$. $G(\varpi, N)$ is time-independent and every row is the same. We will use this cost function for regularising the others and for use on its own for comparison against a 'naive' competitor. Increasing values of ϖ will tighten up the Gaussian although after experimentation we observed that 1 was always the best value and stuck with that.

6.7 Mixing Cost Functions

We mix cost matrices together by adding them. In our experiments we will have a parameter for each cost matrix $\in [0,1]$ to show its contribution to the mixture. The cost matrices will be multiplied by this number before being mixed.

See Figure 4 for an illustration of a single cost matrix and a mixture.

7 Computing Best Segmentation

We obtain the cost of a full segmentation by summing the costs of its tracks. The goal is now to compute efficiently the segmentation of least cost.

We want to reconstruct to find m track boundaries ($m+1$ tracks).

A sequence $\boldsymbol{t} = (t_1, \ldots, t_{m+1})$ is called an m/T-segmentation if

$$1 = t_1 < \ldots < t_m < t_{m+1} = T + 1.$$

m is the number of tracks we are trying to find and is a parameter of the algorithm. We use the interpretation that track $i \in \{1, \ldots, m\}$ comprises times $\{t_i, \ldots, t_{i+1} - 1\}$. Let $\mathbb{S}_m^T$ be the set of all m/T-segmentations. Note that there are a very large number of possible segmentations

$$|\mathbb{S}_m^T| \;=\; \binom{T-1}{m-1} = \frac{(T-1)!}{(m-1)!\,(T-m)!} = \frac{(T-1)(T-2)\cdots(T-m+1)}{(m-1)!} \geq \left(\frac{T}{m}\right)^{m-1}.$$

For large values of T, considering all possible segmentations using brute force is infeasible. For example, a two hour long show with 25 tracks would have more than

$$\left(\frac{60^2 \times 2}{25}\right)^{24} \approx 1.06 \times 10^{59}$$

possible segmentations.

We can reduce this number slightly by imposing upper and lower bounds on the song length. Recall that W is the upper bound (in seconds) of the song length, w the lower bound (in seconds) and m the number of tracks. With the track length restriction in place, the number of possible segmentations is still massive. A number now on the order of 10^{56} for a two hour show with 25 tracks, $w = 190$ and $W = 60 \times 15$.

Let $N(T, W, w, m)$ be the number of segmentations with time T (in tiles),

We can write the recursive relation

$$N(T, W, w, m) = \sum N(t_m - 1, W, w, m-1),$$

where the sum is taken over t_m such that

$$t_m \leq T - w + 1 \qquad t_m \geq T - W + 1$$
$$t_m \geq (m-1)w + 1 \qquad t_m \leq (m-1)W + 1$$

The first two inequalities mean that the length of the last track is within an acceptable boundary between w and W. The last two inequalities mean that the lengths of the first $m-1$ tracks are within the same boundaries.

We calculated the value of $N(\ 7000, 60 \times 15, 190, 25\ )$ and got 5.20×10^{56} which is still infeasible to compute with brute force.

Our solution to this problem is to find a dynamic programming recursion.

The loss of an m/T-segmentation $\boldsymbol{t}$ is

$$\ell(\boldsymbol{t}) \;=\; \sum_{i=1}^{m} C(t_i, t_{i+1} - 1)$$

We want to compute

$$\mathcal{V}_m^T \;=\; \min_{\boldsymbol{t} \in \mathbb{S}_m^T} \ell(\boldsymbol{t})$$

To this end, we write the recurrence

$$\mathcal{V}_1^t \;=\; C(1, t)$$

and for $i \geq 2$

$$\begin{aligned}
\mathcal{V}_i^t &= \min_{\boldsymbol{t} \in \mathbb{S}_i^t} \ell(\boldsymbol{t}) \\
&= \min_{t_i} \min_{\boldsymbol{t} \in \mathbb{S}_{i-1}^{t_i - 1}} \ell(\boldsymbol{t}) + C(t_i, t) \\
&= \min_{t_i} C(t_i, t) + \min_{\boldsymbol{t} \in \mathbb{S}_{i-1}^{t_i - 1}} \ell(\boldsymbol{t}) \\
&= \min_{t_i} C(t_i, t) + \mathcal{V}_{i-1}^{t_i - 1}
\end{aligned}$$

In this formula t_i ranges from $t - W$ to $t - w$. We have $T \times m$ values of $\mathcal{V}_m^T$ and calculating each takes at most $O(W)$ steps. The total time complexity is $O(TWm)$.

8 Confidence Intervals

It may be useful for some applications to build a framework to allow confidence intervals for our predicted indices. This may also be useful for meaningful comparison of cost matrices.

8.1 Posterior Marginal of Song Boundary

Fix a learning rate η, and fix T and m. Let

$$P(j,s) = \frac{\sum_{\boldsymbol{t}\in\mathbb{S}_m^T : t_j = s} e^{-\eta\,\ell(\boldsymbol{t})}}{\sum_{\boldsymbol{t}\in\mathbb{S}_m^T} e^{-\eta\,\ell(\boldsymbol{t})}}$$

That is, $P(j,s)$ is the "posterior probability" that song j starts at time s.

To compute $P(j,s)$, we need an extended notion of segmentation. We call $\boldsymbol{t}$ a $m/F:T$ segmentation if

$$F = t_1 < \ldots < t_m < t_{m+1} = T+1.$$

Let $\mathbb{S}_m^{F:T}$ be the set of all $m/F-T$-segmentations. We have

$$\sum_{\boldsymbol{t}\in\mathbb{S}_m^T : t_j = s} e^{-\eta\,\ell(\boldsymbol{t})} = \sum_{\substack{\boldsymbol{t}\in\mathbb{S}_{j-1}^{s-1},\\ \boldsymbol{t}'\in\mathbb{S}_{m-j+1}^{s:T}}} e^{-\eta(\ell(\boldsymbol{t})+\ell(\boldsymbol{t}'))} = \left(\sum_{\boldsymbol{t}\in\mathbb{S}_{j-1}^{s-1}} e^{-\eta\,\ell(\boldsymbol{t})}\right)\left(\sum_{\boldsymbol{t}\in\mathbb{S}_{m-j+1}^{s:T}} e^{-\eta\,\ell(\boldsymbol{t})}\right)$$

which upon abbreviating

$$\mathcal{H}_m^t = \sum_{\boldsymbol{t}\in\mathbb{S}_m^t} e^{-\eta\,\ell(\boldsymbol{t})} \qquad \mathcal{T}_m^f = \sum_{\boldsymbol{t}\in\mathbb{S}_m^{f:T}} e^{-\eta\,\ell(\boldsymbol{t})}$$

means that we can write

$$P(j,s) = \frac{\mathcal{H}_{j-1}^{s-1}\cdot\mathcal{T}_{m-j+1}^{s}}{\mathcal{H}_m^T}.$$

So it suffices to compute $\mathcal{H}_m^t$ and $\mathcal{T}_m^t$ for all relevant t and m. We use

$$\mathcal{H}_1^t = e^{-\eta C(1,t)} \qquad \mathcal{T}_1^f = e^{-\eta C(f,T-f+1)}$$

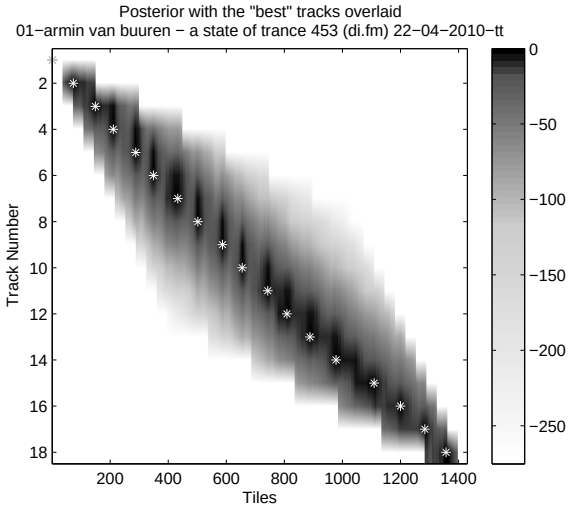


Figure 6: A visualization of $log(P(j,s))$ ($\eta = 10$) for one of the shows in the test set using the cost matrix parameters from experiment 10. The actual tracks are overlaid as white crosses.

and for $m \geq 2$

$$\begin{aligned}
\mathcal{H}_m^t &= \sum_{t_m}\sum_{\boldsymbol{t}\in\mathbb{S}_{m-1}^{t_m-1}} e^{-\eta(\ell(\boldsymbol{t})+C(t_m,t-t_m+1))} \\
&= \sum_{t_m} e^{-\eta C(t_m,t-t_m+1)} \sum_{\boldsymbol{t}\in\mathbb{S}_{m-1}^{t_m-1}} e^{-\eta\,\ell(\boldsymbol{t})} \\
&= \sum_{t_m} e^{-\eta C(t_m,t-t_m+1)}\mathcal{H}_{m-1}^{t_m-1} \\
\mathcal{T}_m^f &= \sum_{t_2}\sum_{\boldsymbol{t}\in\mathbb{S}_{m-1}^{t_2:T}} e^{-\eta(C(f,t_2-f)+\ell(\boldsymbol{t}))} \\
&= \sum_{t_2} e^{-\eta C(f,t_2-f)} \sum_{\boldsymbol{t}\in\mathbb{S}_{m-1}^{t_2:T}} e^{-\eta\,\ell(\boldsymbol{t})} \\
&= \sum_{t_2} e^{-\eta C(f,t_2-f)}\mathcal{T}_{m-1}^{t_2}
\end{aligned}$$

See Figure 6 for an example of the posterior for a radio show.

8.2 Posterior Marginal of Song Position

Fix a learning rate η, and fix T and m. Let

$$P(j,s,f) \;=\; \frac{\sum_{\boldsymbol{t}\in\mathbb{S}_m^T : t_j = s \wedge t_{j+1}-1=f} e^{-\eta\,\ell(\boldsymbol{t})}}{\sum_{\boldsymbol{t}\in\mathbb{S}_m^T} e^{-\eta\,\ell(\boldsymbol{t})}}$$

That is, $P(j,s,f)$ is the "posterior probability" that song j starts at time s and finishes at time f. In the same vein as the last section, we now get

$$P(j,s,f) \;=\; \frac{\mathcal{H}_{j-1}^{s-1}\cdot e^{-\eta C(s,f-s+1)}\cdot \mathcal{T}_{m-j}^{f+1}}{\mathcal{H}_m^T}.$$

8.3 Track Index Confidence

We can use the posterior marginal of song boundary to give estimates of confidence on track index placement and time accuracy.

To estimate the uncertainty of correct track alignment, we select the next highest probability of other track placements at the same time of the optimal placement from $P(j,s)$ and normalize them by the probability of the optimal placement.

Let track index confidence

$$I(j) = 1 - \frac{P(\{1,\ldots,M\}\setminus j, \mathrm{SortInd}(P(j,1,\ldots,T)_1)^2}{\max(P(j,1,\ldots,T)_2)}$$

where $\mathrm{SortInd}(l)$ will return the original indices corresponding to the sorted list of l.

8.4 Track Time Confidence

Track time uncertainty is estimated by normalizing the value of the next most significant peak in $P(j,\forall s)$ by the probability of actual track placement (which is the most significant peak). Let

$$\tilde{I}(j) = 1 - \frac{\mathrm{Peaks}(P(j,1,\ldots,T))_2}{\mathrm{Peaks}(P(j,1,\ldots,T))_1}$$

where $Peaks(s_i)$ is a peak finding algorithm that returns the peaks in descending order of magnitude (we used the `findpeaks` function in MatLab).

See Figure 8 for an illustration of $\tilde{I}(j)$ and $I(j)$.

9 Methodology

We selected 6 shows at random (two of each show type) to use to develop the cost matrices and to find an optimal set of parameters using a random search. These shows were A State of Trance 453 and 462, Magic Island 98 and 112 and Trance Around The World 364 and 372.

In our experiments we decided to fix the tile size at 5 seconds for the sake of speed. A lower tile size does increase accuracy but not significantly. Higher tile sizes can perform more robustly (fewer catastrophic misalignments) but progressively lose out on accuracy.

In our results we will perform some obvious combinations of cost matrices. We also performed a random search of cost matrix contributions and parameters. This has also been provided in the results (see †, experiment 12 in the results). We selected the best set of parameters based on the lowest cost sum of heuristic precision and mean average precision $\sum \hat{A}(P,A) + \hat{H}(P,A) \to \min_i$ for all shows.

10 Materials

All the code presented in this paper with the small working test set is available on GitHub [3]. The large data set ($\approx$ 100GB) we received from Dennis Goncharov can easily be made available on request (it is in a Google Drive account).

11 Results

Please see Table 1 for the main table of results, Figure 7 for a histogram of predicted versus actual differences for experiment 10. Figure 8 shows a visualization of the confidences and relative performance in relation to show progression.

Our parameter search was slightly limited (we searched 2000 random parameters), and should be done again more extensively.

Adding the Gaussian cost matrix in most cases improves performance because it regularises the track

[3] `github.com/ecsplendid/DanceMusicSegmentation`

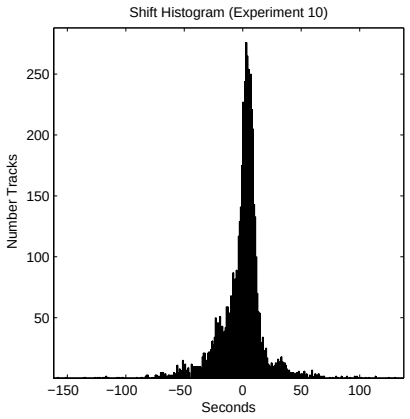


Figure 7: Histogram of the differences between reconstructed and human captured time indices on experiment 10. The slight bumps are misplacements of tracks. The humans are much more likely to place an index before a predicted index.

index estimates (see experiment 6 and the improvement in 11). Using the Gaussian cost matrix on its own makes a useful comparison to how our method in general improves on a potential naive approach to the problem.

The symmetry sum is the best performing single cost matrix with an even incentive bias with summation and evolution close behind. The gains from decreasing window size past $T = 5$ asymptotically decay which leads us to believe we are approaching some kind of limit. More interesting is that when we perform a random parameter search we also hit a barrier when we would have expected further improved results. We estimate that we are now limited by the error variable associated with the human indices more so than our own method.

On our previous work we were using a disincentive only summation matrix, and found that normalizing it on the square root of the width produced the best result. This would have been necessary to encourage placement of longer tracks as no incentive was present. So experiment 11 is roughly comparable and indeed produces the same overall mean average to that previous experiment ($\approx$ 20S). Note that we no longer discard any shows from evaluation which makes the result stronger.

Figure 8 is interesting. It seems to suggest that with the summation matrix alone (and we suspect all of the matrices when used in isolation), we suffered significant loss in confidence and performance in the middle of the shows. This would imply that our method was failing to some extent, and to make matters worse; the magnitude of failure depended on the number of tracks. Apparently the mixture of cost functions removes this effect.

12 Summary

We believe our algorithm would be useful for segmenting DJ-mixed audio streams in batch mode. It would be excellent if Spotify[4] for example started to do something similar. Spotify is an on-line music service with many electronic dance music radio shows with the track listing in text. This method would allow them to reliably segment the shows, and they could display an interactive segmentation in the music player.

The new cost matrices in combination improve robustness significantly over single matrices or regularised single matrices (as in our last paper). We are seeing about a 50% improvement in overall mean accuracy over single cost matrices that are correctly normalized. The new cost matrices improve on many drawbacks of our previous work (mainly that it was vulnerable to dissimilar regions within tracks).

Our method still has one key drawback that we are aware of. This is the rare instance where there are head or tail segments to a track that seem independent from the rest of the track. When these are small they usually get absorbed without any problems but they can cause misplacements. In spite of this issue we suspect that our predictions are more accurate and more consistent than the human equivalents while not being as precise in situations when our index agrees with theirs.

We would also like to implement some of the methods in the literature (which were mostly designed for scene analysis) to see if we outperform them. It

[4] `http://spotify.com`

would be tricky to get an exact comparison because we could not find an unsupervised deterministic algorithm which finds a fixed number of strictly contiguous clusters. We could however adapt existing algorithms to get a like for like comparison. We would like to evaluate the performance of J Theiler's contiguous K-means algorithm in particular [17] and also similar algorithms. We have the property of being deterministic but probabilistic methods should be explored. Theiler's algorithm would require some modification to work in this scenario because we require strictly contiguous clusters, not just a contiguity bias.

13 Acknowledgements

We would like to thank Mikael Lindgren and Dennis Goncharov from cuenation[5] for their help explaining how they make cue sheets and for providing the data set to test the algorithm on.

References

[1] J. Matejka, T. Grossman, and G. Fitzmaurice, "Swifter: Improved online video scrubbing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, (New York, NY, USA), pp. 1159–1168, ACM, 2013.

[2] G. Tzanetakis and F. Cook, "A framework for audio analysis based on classification and temporal segmentation," in *EUROMICRO Conference, 1999. Proceedings. 25th*, vol. 2, pp. 61–67, IEEE, 1999.

[3] J. Foote, "Visualizing music and audio using self-similarity," in *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pp. 77–80, ACM, 1999.

[4] J. Foote, "A similarity measure for automatic audio classification," in *Proc. AAAI 1997 Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora*, 1997.

[5] `http://www.cuenation.com`

[5] J. Foote, "Automatic audio segmentation using a measure of audio novelty," in *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, vol. 1, pp. 452–455, IEEE, 2000.

[6] J. T. Foote and M. L. Cooper, "Media segmentation using self-similarity decomposition," in *Electronic Imaging 2003*, pp. 167–175, International Society for Optics and Photonics, 2003.

[7] J. Foote and M. Cooper, "Visualizing musical structure and rhythm via self-similarity," in *Proceedings of the 2001 International Computer Music Conference*, pp. 419–422, 2001.

[8] M. M. Goodwin and J. Laroche, "Audio segmentation by feature-space clustering using linear discriminant analysis and dynamic programming," in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pp. 131–134, IEEE, 2003.

[9] M. M. Goodwin and J. Laroche, "A dynamic programming approach to audio segmentation and speech/music discrimination," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 4, pp. iv–309, IEEE, 2004.

[10] G. Peeters, A. La Burthe, and X. Rodet, "Toward automatic music audio summary generation from signal analysis," in *Proc. of ISMIR*, pp. 94–100, 2002.

[11] G. Peeters, "Deriving musical structures from signal analysis for music audio summary generation:"sequence" and "state" approach," *Computer Music Modeling and Retrieval*, pp. 169–185, 2004.

[12] E. Peiszer, T. Lidy, and A. Rauber, "Automatic audio segmentation: Segment boundary and structure detection in popular music," *Proc. of LSAS*, 2008.

[13] T. Scarfe, W. M. Koolen, and Y. Kalnishkan, "A long-range self-similarity approach to segmenting dj mixed music streams," in *Artificial Intel-*

ligence Applications and Innovations, pp. 235–244, Springer, 2013.

[14] H. Nyquist, "Certain topics in telegraph transmission theory," *American Institute of Electrical Engineers, Transactions of the*, vol. 47, no. 2, pp. 617–644, 1928.

[15] M. Frigo and S. G. Johnson, "The fftw web page," 2004.

[16] G. Tzanetakis and P. Cook, "Multifeature audio segmentation for browsing and annotation," in *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pp. 103–106, IEEE, 1999.

[17] J. P. Theiler and G. Gisler, "Contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation," in *Optical Science, Engineering and Instrumentation'97*, pp. 108–118, International Society for Optics and Photonics, 1997.

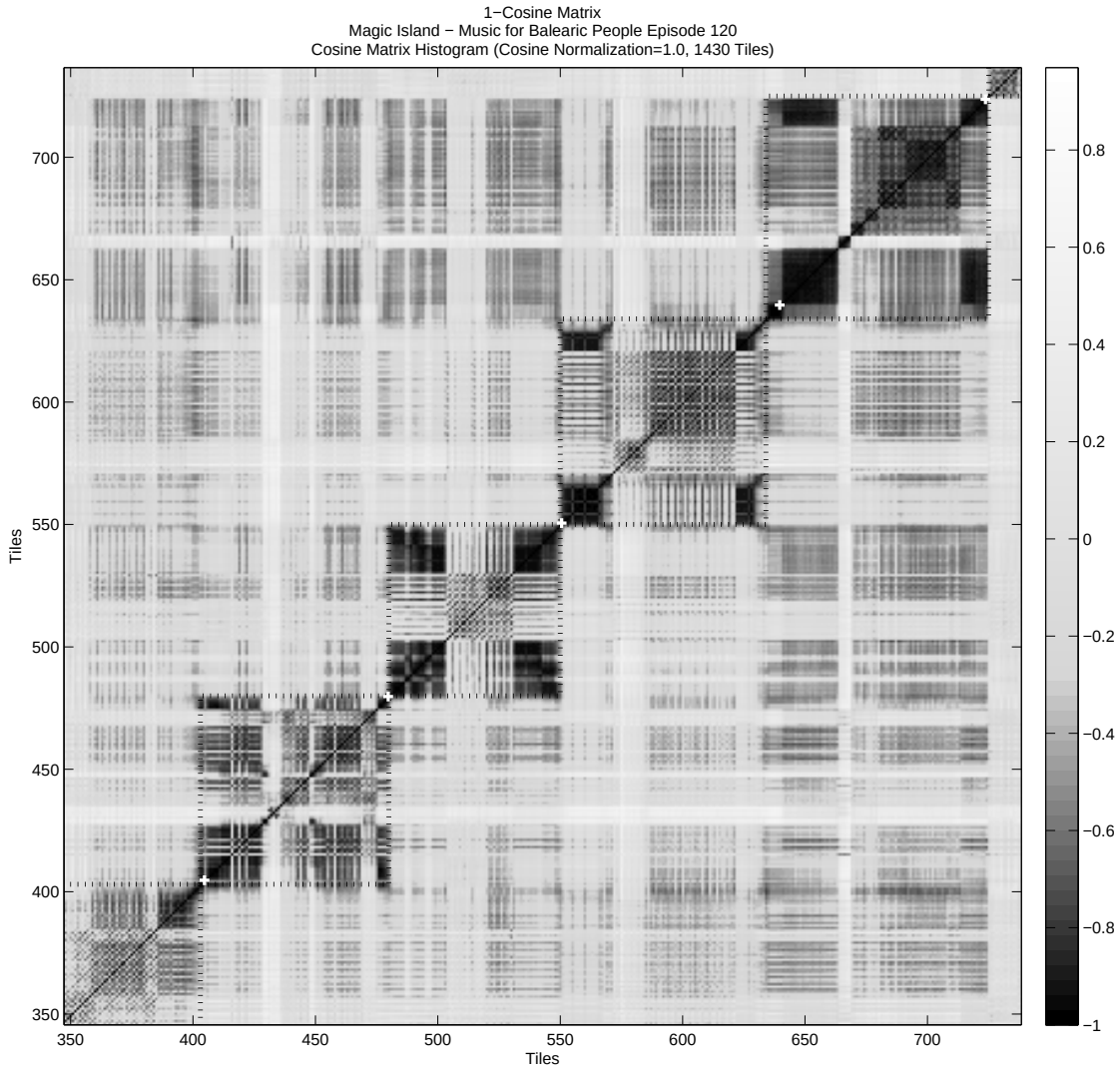


Figure 3: An illustration of the similarity matrix S with the actual indices drawn on with white crosses, and our reconstructed indices indicated with the black dotted lines. There are examples here of evolutionary repetition ($t = 500, \ldots, 550$), static contiguity everywhere where there is solid black, and symmetry on the middle two tracks.

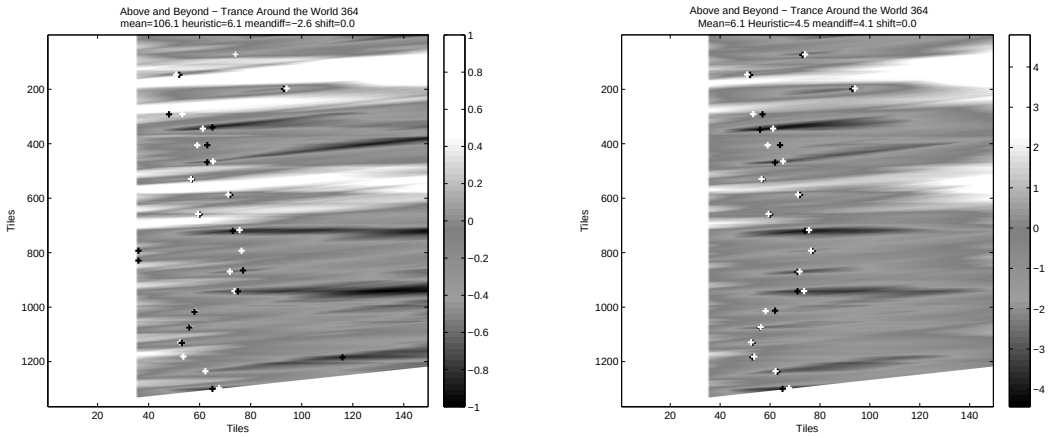


Figure 4: Cost matrices for Magic Island episode 110. On the left is the cost matrix parameters from experiment 5 (summation only with incentive bias 0.5) and the right the mixed cost matrix using parameters from experiment 10. On the right, the predicted (white) and actual tracks (black) are shown in place. Note that the white space is infinite corresponding to the minimum and maximum track parameters w and W.

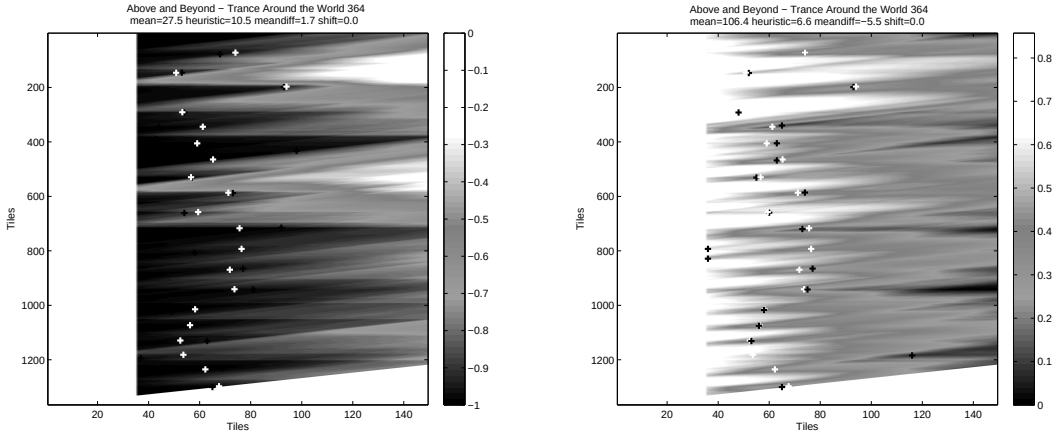


Figure 5: Summation cost matrices for Magic Island episode 110. The matrix on the left has an incentive bias $\Omega = 1$ and therefore only contains disincentives. However the matrix on the right is the other extreme, $\Omega = 1$ containing only incentives.

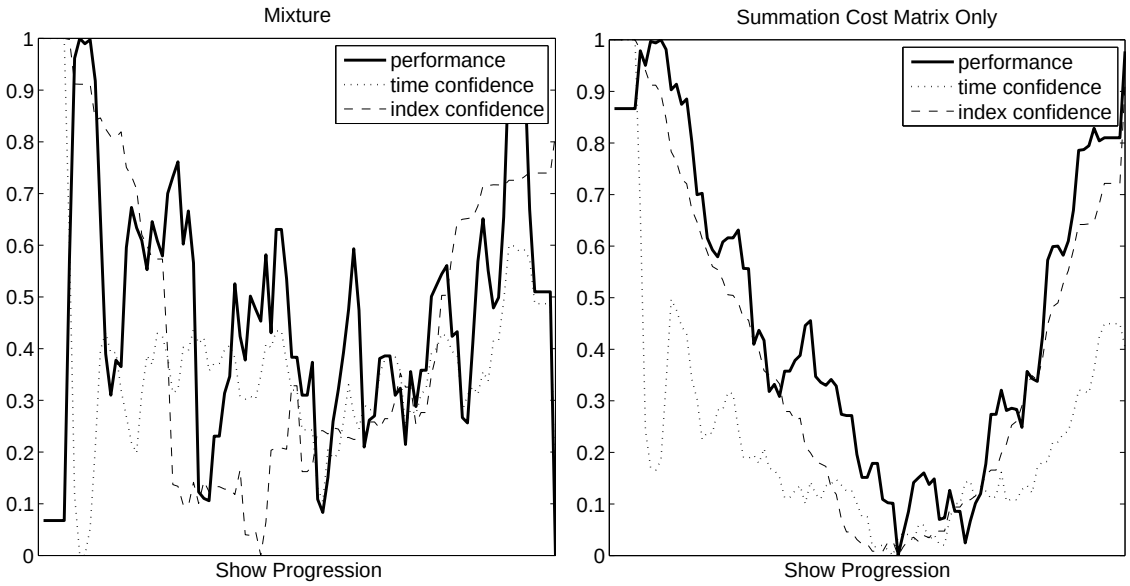


Figure 8: The relative performance and confidence ($\eta = 10$) of two cost matrices. The left is experiment 10 which has a mixture of cost matrices (see the main Results table) and the right is experiment 5 which is only the summation matrix. Ostensibly the summation cost matrix struggles towards the middle of the shows.

Table 1: Main results. Tile size T was 5s for all results. Shift refers to the mean average of the differences in predictions versus human captured indices. $\star$; $h = 1662$Hz, $b = 5$Hz, $\hat{c} = 1$, $\rho = 20$ †(*randomsearch*); $h = 1662$Hz, $b = 5$Hz, $\hat{c} = 1.2$, $\rho = 26$. Blank cells indicate no contribution from that cost matrix.

Experiment	h, l, b, $\hat{c}$, ρ	Sym Sum	Sym Diff	Symmetry	Contig Past	Contig Fut.	Summation	Gaussian	Evolution	Mean (S)	Heuristic (S)	Shift (S)	60s(%)	30s(%)	20s(%)	10s(%)	5s(%)	1s(%)
1	*	$1.0_{\Omega^{\frac{1}{2}}}$								32.13	8.59	1.37	96.06	87.28	78.15	60.48	33.82	20.89
2	*		$1.0_{\Omega^{\frac{1}{2}}}$							35.20	16.56	2.52	91.83	73.00	58.54	36.40	19.75	12.41
3	*			$1.0_{\Omega^{\frac{1}{2}}}$						68.33	11.50	1.39	92.70	80.35	69.51	50.59	27.55	17.62
4	*				$1.0_{\Omega^{\frac{1}{2}}}$					152.58	21.75	5.36	79.64	63.05	49.20	33.12	17.93	10.90
5	*					$1.0_{\Omega^{\frac{1}{2}}}$				160.09	27.60	12.25	76.97	54.36	42.27	27.43	14.67	9.04
6	*						$1.0_{\Omega^{\frac{1}{2}}}$			34.30	8.65	1.52	95.85	87.17	77.76	60.48	33.66	20.81
7	*							$1.0_{\Omega^{\frac{1}{2}}}$		97.83	73.16	−6.08	43.37	22.89	15.53	7.54	3.86	2.38
8	*	$1.0_{\Omega^{\frac{1}{2}}}$							$1.0_{\Omega^{\frac{1}{2}}}$	33.99	8.49	3.95	96.01	87.76	79.61	60.70	33.82	21.40
9	*	$1.0_{\Omega^{\frac{1}{2}}}$	$1.0_{\Omega^{\frac{1}{2}}}$						$1.0_{\Omega^{\frac{1}{2}}}$	24.88	8.43	3.01	97.20	89.00	80.24	60.55	33.57	20.98
10	*	$1.0_{\Omega^{\frac{1}{2}}}$	$1.0_{\Omega^{\frac{1}{2}}}$	$1.0_{\Omega^{\frac{1}{2}}}$						16.31	8.09	2.40	97.61	89.23	80.93	61.87	34.85	22.07
11	*						$1.0_{\Omega^{0.8}}$	$1.0_{\Omega^{\frac{1}{2}}}$		21.57	8.92	1.52	97.19	87.57	78.19	59.34	32.65	20.64
12	†	$0.4_{\Omega^{0.4}}$	$0.1_{\Omega^{0.1}}$	$0.1_{\Omega^{0.7}}$	$0.2_{\Omega^{0.2}}$	$0.0_{\Omega^{0.7}}$	$0.6_{\Omega^{0.6}}$	$0.7_{\Omega^{0.8}}$	$0.4_{\Omega^{0.1}}$	16.91	8.27	2.33	97.93	89.76	80.74	61.22	34.65	22.16