Grateful Live: Mixing Multiple Recordings of a Dead Performance into an Immersive Experience

Wilmering, T; Thalmann, F; Sandler, MB

Audio Engineering Society

Convention Paper 9614

Presented at the 141st Convention
2016 September 29 – October 2, Los Angeles, CA, USA

Grateful Live: Mixing Multiple Recordings of a Dead Performance into an Immersive Experience

Thomas Wilmering, Florian Thalmann, and Mark B. Sandler

Centre for Digital Music (C4DM), Queen Mary University of London, London E1 4NS, UK

Correspondence should be addressed to Thomas Wilmering (`t.wilmering@qmul.ac.uk`)

ABSTRACT

Recordings of historical live music performances often exist in several versions, either recorded from the mixing desk, on stage, or by audience members. These recordings highlight different aspects of the performance, but they also typically vary in recording quality, playback speed, and segmentation. We present a system that automatically aligns and clusters live music recordings based on various audio characteristics and editorial metadata. The system creates an immersive virtual space that can be imported into a multichannel web or mobile application allowing listeners to navigate the space using interface controls or mobile device sensors. We evaluate our system with recordings of different lineages from the Internet Archive Grateful Dead collection.

1 Introduction

Recordings of historical live music performances often exist in several versions, either recorded from the mixing desk, on stage, or by audience members from various positions in the performance space. These recordings, both soundboard recordings and audience-made recordings, highlight different aspects of the performance, but they also typically vary in recording quality, playback speed, and segmentation. In this paper we present a system that automatically aligns and clusters live music recordings based on various audio characteristics and editorial metadata. The system creates an immersive virtual space that can be imported into a multichannel web or mobile application where listeners can navigate it using interface controls or mobile device sensors. We evaluate our system with items from the Internet Archive Grateful Dead collection[1], which contains recordings with many different lineages of a large number of performances. The research is motivated by the continuing interest in the Grateful Dead and their performances, evidenced by the large amount of information available in the literature and on the Web [1].

We first describe the content of the Internet Archive Grateful Dead Collection, before discussing concert recording lineages and the strategy for choosing the material for this study. This is followed by a brief discussion of the audio feature extraction performed on the collection. After describing and evaluating the algorithms employed in the analysis and clustering of the audio material, we draw conclusions and outline

[1] https://archive.org/details/GratefulDead

future work.

2 The Internet Archive Grateful Dead Collection

The Live Music Archive (LMA)[2], part of the Internet Archive, is a growing openly available collection of over 100,000 live recordings of concerts, mainly in rock genres. Each recording is accompanied by basic unstructured metadata describing information including dates, venues, set lists and the source of the audio files. The Grateful Dead collection is a separated collection, created in 2004, consisting of both audience-made and soundboard recordings of Grateful Dead concerts. Audience-made concert recordings are available as downloads while soundboard recordings are accessible to the public in streaming format only.

2.1 Recording Lineages

A large number of shows is available in multiple versions. At the time of writing the Grateful Dead collection consisted of 10537 items, recorded on 2024 dates. The late 1960s saw a rise in fan-made recordings of Grateful Dead shows by so-called *tapers*. Indeed, the band encouraged the recording of there concerts for non-commercial use, in many cases providing limited dedicated *taper tickets* for their shows. The Tapers set up their equipment in the audience space, typically consisting of portable, battery-powered equipment including a cassette or DAT recorder, condenser microphones, and microphone preamplifiers. Taping and trading of Grateful Dead shows evolved into a subculture with its own terminology and etiquette [2]. The Internet Archive Grateful Dead collection consists of digital transfers of such recordings. Their sources can be categorised into three main types [3]:

Soundboard (SBD) – Recordings made from the direct outputs of the soundboard at a show, which usually sound very clear with no or little crowd noise. Cassette SBDs are sometimes referred to as SBDMC, DAT SBDs as SABD or DSB. There have been instances where tapes made from monitor mixes have been incorrectly labelled as SBD.

Audience (AUD) – Recordings made with microphones in the venue, therefore including crowd noise. These are rarely as clean as SBD. At Grateful Dead shows the *taper section* for taper ticket holders was located behind the soundboard. Recordings at other locations may be labelled, for instance, a recording taped in front of the soundboard may labeled FOB (front of board).

Matrix (MAT) – Recordings produced by mixing two or more sources. These are often produced by mixing an SBD recording with AUD recordings, therefore including some crowd noise, while preserving the clean sound of the SBD. The sources for the matrix mixes are usually also available separately in the archive.

Missing parts in the recordings, resulting for example from changing of the tape, are often patched with material from other recordings in order to produce a complete, gapless recording. The Internet Archive Grateful Dead Collection's metadata provides separate entries for the name the taper and the name of the person who transferred the tape into the digital domain (*transferer*). Moreover, the metadata includes additional editorial, unstructured metadata about the recordings. In addition to the concert date, venue and playlist, the lineage of the archive item is given with varying levels of accuracy. For instance, the lineage of the recording found in archive item*gd1983-10-15.110904.Sennheiser421-daweez.D5scott.flac16* is described as:

Source: 2 Sennheiser 421 microphones (12th row-center) → Sony TC-D5M - master analog cassettes

Lineage: Sony TC-D5M (original record deck) → PreSonus Inspire GT → Sound Forge → .wav files → Trader's Little Helper → flac files

Source describes the equipment used in the recording, *Lineage* the lineage of the digitisation process. The above example describes a recording produced with two dynamic cardioid microphones and recorded with a portable cassette recorder from the early 1980s. The *lineage* metadata lists the playback device, and the audio hardware and software used to produce the final audio files. Each recording in the collection is provided as separated files reflecting the playlist. The segment boundaries for the files in different versions of one concert differ, since the beginning and end of songs in a live concert are not often clear. Moreover, the way the time between songs, often filled with spoken voice or instrument tuning, is handled differently. Some versions include separate audio files for these sections, while in other versions it may be included in the previous or following track.

[2] https://archive.org/details/etree/

2.2 Choice of Material for this Study

By analysing the collection we identified concerts available in up to 22 versions. Figure 1 shows the number of concerts per year, and the average number of different versions of concerts for each year. On average, there are 5.2 recordings per concert available. For the material for this study we chose 2 songs from 8 concerts each. The concerts were selected from those having the highest number of versions in the collection, all recorded at various venues in the USA between 1982 and 1990. Many versions partially share the lineage or are derived from mixing several source in a matrix mix. In some cases recordings only differ in the sampling rate applied in the digitisation of the analog source[3]. Table 1 shows the concert dates selected for the study, along with the available recording types and number of distinct tapers and transferers, identified by analysing the editorial metadata. We excluded surround mixes, which are typically derived from sources available separately in the collection.

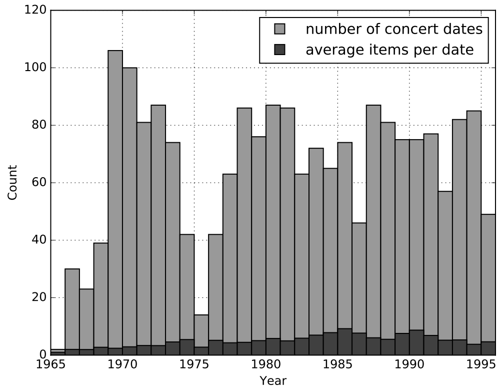

Fig. 1: Number of concert dates per year and the average number of versions per concert per year in the Internet Archive Grateful Dead collection.

Concert date	Items	Tapers	Transferers	Soundboard	Audience	Matrix
1982-10-10	22	5	7	6	12	4
1983-10-15	18	11	10	2	15	1
1985-07-01	17	8	12	4	12	1
1987-09-18	19	5	12	8	5	6
1989-10-09	16	5	9	4	10	2
1990-03-28	19	8	15	7	10	2
1990-03-29	19	7	11	5	12	2
1990-07-14	18	7	10	6	10	2

Table 1: Number of recordings per concert used in the experiments. Source information and the number of different tapers and transferers are taken from editorial metadata in the Internet Archive.

3 Feature Extraction

In earlier work a linked data service that publishes the previously unstructured metadata from the LMA has been created [4]. Within the CALMA (Computational Analysis of the Live Music Archive) project [5, 6] we developed tools to supplement this performance metadata with automated computational analysis of the audio data using Vamp feature extraction plugins[4]. This set of audio features includes high level descriptors such as chord times and song tempo, as well as lower level features. Among them chroma features [7] and MFCCs [8], which are of particular interest of this study and have been used for measuring audio similarity [9, 10]. A chromagram describes the spectral energy of the 12 pitch classes of an octave by quantising the frequencies of the spectral analysis resulting in a 12 element vector. It can be defined as an octave-invariant spectrogram taking into account aspects of musical perception. MFCCs include a conversion of Fourier coefficients to the Mel-scale and represent the spectral envelope of a signal. They have originally been used in automatic speech recognition. For an extensive overview audio features in the context of content-based audio retrieval see [11].

4 Creating the Immersive Experience

As a first step towards creating a novel browsing and listening experience, we consider all sets of recordings of single performances, which are particularly numerous in the Grateful Dead collection of the Internet Archive (see Section 2). Our goal is to create an immersive space using binaural audio techniques, in which the

[3] The collection includes audio files with sample rates of 44.1kHz, 48kHz and 96kHz

[4] http://www.vamp-plugins.org

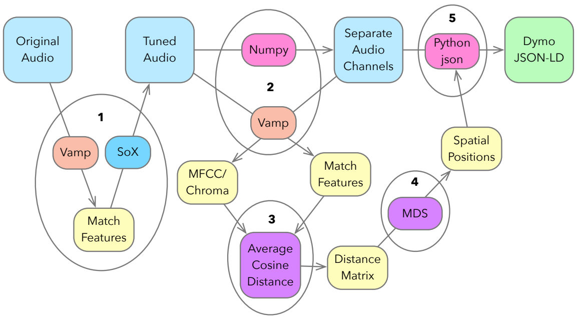

Fig. 2: The five-step procedure for creating an immersive experience.

single channels of all recordings are played back synchronously and which listeners can navigate to discover the different characteristics of the recordings. For this, we designed a five-step automated procedure using various tools and frameworks and orchestrated by a Python program[5].

We first align and resample the various recordings, which may be out of synchronisation due to varying tape speeds anywhere in their lineageThen, we align the resampled recordings and extract all features necessary in the later process. We then calculate the average distance for each pair of recordings based on the distances of features in multiple short segments, resulting in a distance matrix for each concert. Next, we perform Multidimensional Scaling (MDS) to obtain a two- or three-dimensional spatial distribution that roughly maintains the calculated average distances between the recordings. Finally, we render the performance binaurally by placing the recordings as sound sources in an immersive space where the listeners can change their position and orientation dynamically. Figure 2 visualises the entire procedure.

4.1 First Alignment, Tuning, and Normalization

Since Vamp Plugins (see Section 3) can deliver their output as linked data we decided to use them in as many steps of the process as possible. The plugins include a tool for alignment of different recordings, the MATCH Plugin[6] which proved to suit our purpose although its algorithm is not optimised for our use case. The MATCH (Music Alignment Tool CHest) [12] is based on an online dynamic time-warping algorithm and specialises in the alignment of recordings of different performances of the same musical material, such as differing interpretations of a work of classical music. Even though it is determined to detect more dramatic tempo changes and agogics, it proved to be well-suited for our recordings of different lineages, most of which exhibit only negligible tempo changes due to uneven tape speed but greater differences in overall tape speed and especially timbre, which the plugin manages to align.

In our alignment process we first select a reference recording a, usually the longest recording when aligning differently segmented songs, or the most complete when aligning an entire concert. Then, we extract the MATCH `a_b` featuresfor all other recordings $b^1 ... b^n$ with respect to the reference recording a The MATCH features are represented as a sequence of time points b_i^j in recording b^j with their corresponding points in a, $a_i^j = f_{b^j}(b_i^j)$. With the plugin's standard parameter configurationthe step size between time points b_k^j and b_{k+1}^j is 20 milliseconds. Based on these results we select an early time point in the reference recording a_e that has a corresponding point in all other recordings as well as a late time point a_l with the same characteristics. From this we determine the playback speed ratio γ^j of each b^j relative to a as follows:

$$\gamma^j = \frac{a_l - a_e}{f_{b^j}^{-1}(a_l) - f_{b^j}^{-1}(a_e)} \quad \text{for } j \in \{1, ..., n\}$$

Using these ratios we then adjust all recordings b^j using the `speed` effect of the SoX command line tool[7] so that their average playback speed and tuning matches the reference recording a. With the same tool we also normalise the tuned recordings before the next feature extraction to ensure that they all have comparable average power, which is significant for adjusting the individual playback levels of the recordings in the resulting immersive experience.

4.2 Second Alignment and Feature Extraction

After the initial aligning and resampling we re-extract the MATCH `a_b` features in order to deal with smaller temporal fluctuations. We then separate all stereo recordings into their individual channels, and cluster each channel separately. This is followed by extracting all features necessary for calculating the distances

[5] https://github.com/florianthalmann/live-cluster

[6] https://code.soundsoftware.ac.uk/projects/match-vamp

[7] http://sox.sourceforge.net

that form the basis for the clustering, typically simply *MFCC* and *Chroma*, for each of the individual channels. If all recordings are stereo, which is usually the case, we obtain n a_b feature files and $2*(n+1)$ files for all other features.

4.3 Calculating the Pairwise Average Distance

A common and simple way to calculate distances between a number of audio files or segments is to create feature vectors for each of them and calculate the distances between these vectors. These feature vectors can be obtained by averaging a number of relevant temporal features over the entire duration of the files. However, even though this way we might get a good idea of the overall sonic qualities of the files, we may ignore local temporal differences which are particularly pronounced in our case, where the position within the audience and from the speakers might create dramatic differences between the recordings. Therefore, instead of simply creating summarising feature vectors, we generated vectors for shorter synchronous time segments throughout the recordings, calculate distances between those, and finally average all pairwise distances between the recordings thus obtained. More specifically, we choose a segment duration d and a number of segments m and we define the following starting points of segments in a:

$$s_k = a_e + k*(a_l - a_e)/m \text{ for } k = 0,\dots,m-1$$

From these, we obtain the following segments in a

$$S_k^a = [s_k, s_k + d]$$

as well as all other recordings b^j which are identical for all of their channels:

$$S_k^{b^j} = [f_{b^j}'^{-1}(s_k), f_{b^j}'^{-1}(s_k + d)]$$

where f'_{b^j} is the assignment function for b^j resulting from the second alignment. We then calculate the normalised averages and variances of the features for each segment $S_k^{b^j}$ which results in m feature vectors v_k^r for each recording for $r \in \{a, b^1, \dots, b^n\}$. Figure 3 shows an example of such feature vectors for a set of recordings of *Looks Like Rain* on October 10, 1982. This example shows the large local differences resulting from different recording positions and lineages. For comparison, Figure 4 shows how the feature vectors averaged over the whole duration of the recordings are much less diverse.

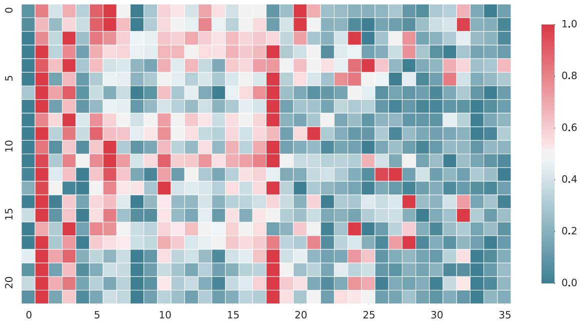

Fig. 3: Averages (18 left columns) and variances (18 right columns) of MFCC features across a 0.5 second segment. Each row is a different channel of a recording of *Looks Like Rain* on October 10, 1982.

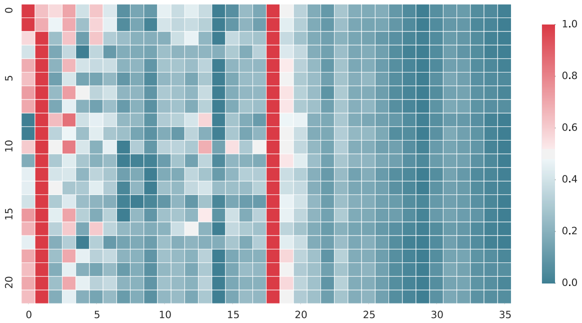

Fig. 4: Averages and variances of MFCC features across a 490 second segment of *Looks Like Rain* on October 10, 1982.

With these feature vectors, we determine the pairwise distance between the recordings a and b^j for each $k = 0,\dots,m$, in our case using the cosine distance, or inverse cosine similarity [9]:

$$d_k(x,y) = 1 - \frac{v_k^x \cdot v_k^y}{||v_k^x|| \cdot ||v_k^y||}$$

for $x, y \in \{a, b^1, \dots, b^n\}$. We then take the average distance

$$d(x,y) = \frac{1}{m} \cdot \sum_k d_k(x,y)$$

for each pair of recordings x, y which results in a distance matrix D such as the one shown in Figure 5. In this matrix we can detect many characteristics of the recordings. For instance, the two channels of recording 1 (square formed by the intersection of rows 3 and 4 and columns 3 and 4) are more distant from each

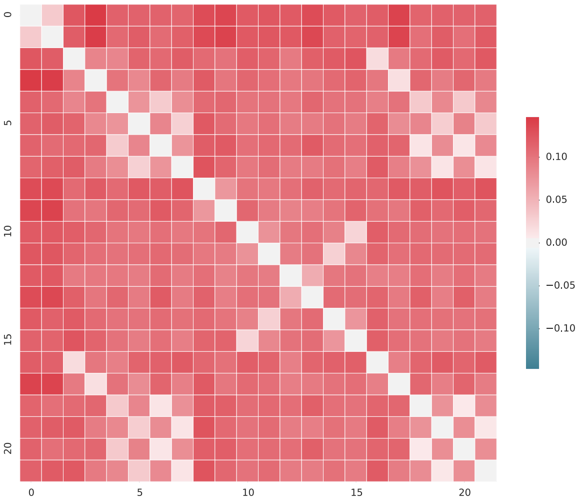

Fig. 5: A distance matrix for the separate channels of 11 different recordings of *Looks Like Rain* on October 10, 1982 (duplicates and conversions were removed). Each pair of rows and pair of columns belong to a stereo recording.

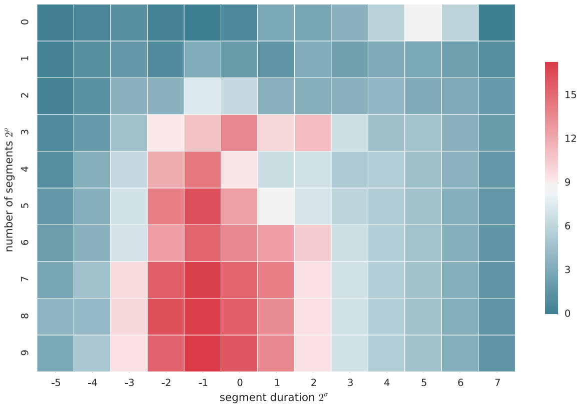

Fig. 6: Values for $eval(D)$ for different combinations of parameters m and d.

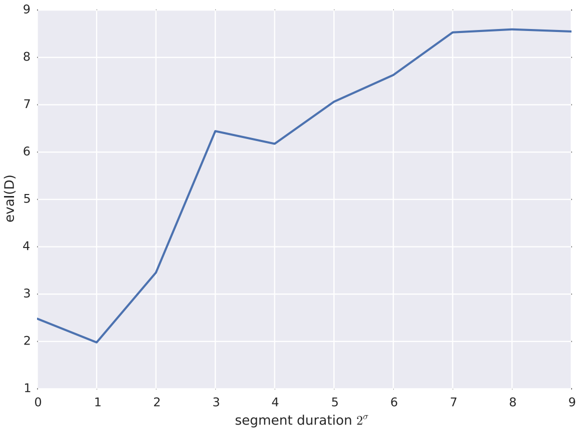

Fig. 7: Values for $eval(D)$ for different m.

other than the two channels of the reference recording 0, which hints at a heightened stereo effect. Recordings 2, 3, 9, and 10 seem to be based on each other where recording 2 seems slightly more distant from the others. Recordings 5 and 7 again see to be based on each other, however, their channels seem to have been flipped.

4.3.1 Determining the Optimal Parameters

In order to improve the final clustering for our purposes we experimented with different parameters m and d in search of an optimal distance matrix. The characteristics we were looking for was a distribution of the distances that includes a large amount of both very short and very long distances, provides a higher resolution for shorter distances, and includes distances that are close to 0. For this we introduced an evaluation measure for the distance distributions based on *kurtosis*, *left-skewedness*, as well as the position of the 5th percentile:

$$eval(D) = \frac{Kurt[D](1 - Skew[D])}{P_5[D]}$$

where *Kurt* is the function calculating the fourth standardised moment, *Skew* the third standardised moment, and P_5 the fifth percentile.[8]

Figure 6 shows a representation of $eval(D)$ for a number of parameter values $m = 2^\rho, d = 2^\sigma$ for $\rho \in \{0, 1, \ldots, 9\}$ and $\sigma \in \{-5, -4, \ldots, 7\}$ and Figures 7 and 8 show the sums across the rows and columns. We see that with increasing m we get more favourable distance distributions and a plateau after about $m = 128$, and an optimal segment length of about $d = 0.5sec$. These are the parameter values we chose for the subsequent calculations yielding satisfactory results. The distance matrix in Figure 5 shows all of the characteristics described above.

[8] We use SciPy and NumPy (http://www.scipy.org) for the calculations, more specifically *scipy.stats.kurtosis*, *scipy.stats.skew*, and *numpy.percentile*.

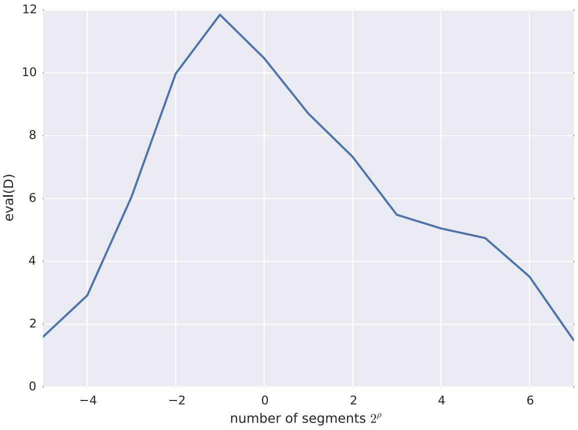

Fig. 8: Values for $eval(D)$ for different d.

4.4 Multidimensional Scaling

After initial experiments with various clustering methods we decided to use metric Multidimensional Scaling (MDS) [13] to create a spatialisation of the different recording of a performance. Metric MDS takes a distance matrix as an input and iteratively finds a spatial arrangement of objects, keeping the distances between the objects as proportional as possible to the distances in a given matrix. This is achieved by minimising a cost function $stress$. We use the implementation available in *scikit learn*, `sklearn.manifold.MDS`[9] with default options to create a two- or three-dimensional distribution of all channels of the given recordings.

Figure 9 shows the two-dimensional spatial arrangement resulting from the distance matrix in Figure 5. The positions of the individual channels closely reflect the characteristics of the recordings we observed in the distance matrix and discussed in Section 4.3, e.g. the proximity of the individual channels of recordings 2, 3, 9, and 10, the flipped channels of 5 and 7, and the lesser stereo effect of 0. In addition, we can also observe relationships that are less visible in the distance matrix, such as the fact that recording 4 is closer to 5 than 7, or even 1.

4.5 Creating Dynamic Music Objects

In our final step, we represent the obtained spatial distribution in a way that is understood by our prototypical applications. We chose to use Dynamic Music Objects (dymos), a music format based on Semantic Web technologies that can be used to create a variety of musical experiences of adaptive, interactive, or otherwise dynamic nature [14]. Dymos are based on the abstract multi-hierarchical music representation system CHARM [15]. A multitude of musical structures can be described this way such as multi-level segmentations, audio processing chains and groups, or spatial arrangements. These structures can be annotated with semantic information extracted from the audio files which will then inform the way they are navigated and played back. On top of this, one can define modifiable musical parameters and their interrelationships [14] and map the signals of various controls, including sensors, UI elements, and auto-controls, to these parameters via arbitrary mapping functions. Dymos can be built into any web or mobile application but there is also a mobile app framework, the *Semantic Music Player*, which can be used to test and distribute specific experiences [16].

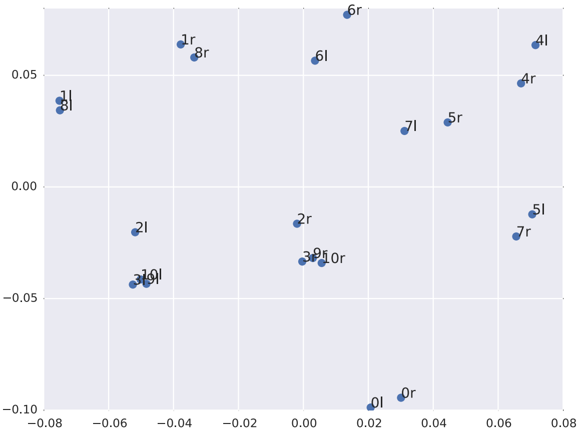

Fig. 9: The two-dimensional MDS cluster resulting from the distances shown in Figure 5. The numbers correspond to the row and column numbers in the matrix and l and r indicate the channel.

The immersive experience described in this paper can easily be built with dymos, by simply taking the positions output of the MDS in the previous step and scaling the positions to an appropriately sized virtual space. We provide a Python script that outputs a JSON-LD representation of a dymo hierarchy with objects for each channel at its corresponding position, as well as the mappings necessary to navigate the space. At this stage, we suggest two interaction schemes. The first

[9] http://scikit-learn.org/stable/modules/generated/sklearn.manifold.MDS.html

one is more traditional, where the users see a graphical representation of the spatial arrangement (similar to Figure 9) and can change their position and orientation with simple mouse-clicks or taps. The second interaction scheme is optimised for mobile devices and maps geolocation sensor inputs to spatial location and compass input to listener orientation. In this way, the listeners can change their position and orientation in the virtual space by physically walking around while listening to a binaural rendering of the music on headphones. Listing 1 illustrates how a rendering definition for the latter interaction scheme looks in JSON-LD.

```
"@context":"http://tiny.cc/dymo-context",
"@id":"deadLiveRendering",
"@type":"Rendering",
"dymo":"deadLiveDymo",
"mappings":[
  {
    "domainDims":[
      {"name":"lat","@type":"GeolocationLatitude"}
    ],
    "function":{"args":["a"],
      "body":"return (a-53.75)/0.2;"},
    "dymos":{"args":["d"],
      "body":"return d.getLevel() == 1;"},
    "parameter":"Distance"
  },
  {
    "domainDims":[
      {"name":"lon","@type":"GeolocationLongitude"}
    ],
    "function":{"args":["a"],
      "body":"return (a+0.03)/0.1;"},
    "dymos":{"args":["d"],
      "body":"return d.getLevel() == 1;"},
    "parameter":"Pan"
  },
  {
    "domainDims":[
      {"name":"com","@type":"CompassHeading"}
    ],
    "function":{"args":["a"],
      "body":"return a/360;"},
    "parameter":"ListenerOrientation"
  }
]
```

Listing 1: A rendering for a localised immersive experience.

5 Results

We take two steps to preliminarily evaluate the results of this study. First, we discuss how the clustering obtained for the test performance *Looks Like Rain* on October 10, 1982 (Section 4) compares to the manual annotations retrieved from the Live Music Archive (Section 2). Then we compare the average distances obtained for different recording types of the entire material selected for this study (Section 2.2).

For the test performance, we already removed duplicates and exact conversions manually, based on file names and annotations. Nevertheless, as described in Sections 4.3 and 4.4, we detect high similarities between recordings 2, 3, 9, and 10. Consulting the annotations we find out that 3, 9, and 10 are all soundboard recordings (SBD) with slightly different lineages. 2 is a matrix recording (MAT) combining the SBD with unknown audience recordings (AUD), which explains the slight difference from the other three which we observed. From the clustering we could hypothesise that one of the AUD used are 1 and 8 based on the direction of 2's deviation. 1 and 8 are both based on an AUD by Rango Keshavan, with differing lineages. 7 is an AUD taped by Bob Wagner and 5 is annotated as by an anonymous taper. We could infer that either 5 was derived from 7 at some point and the taper was forgotten, or it was recorded at a very similar position in the audience. 0 is again by another taper, David Gans, who possibly used a more close microphone setup. 4 and 6 are two more AUDs by the tapers Richie Stankiewicz and N. Hoey. Even though the positions obtained via MDS are not directly related to the tapers' locations in the audience, we can make some hypotheses.

Table 2 presents average distances for the 16 recordings chosen for this study. We assigned a category (AUD, MAT, SBD) to each version based on the manual annotations in the collection (Table 1). We averaged the distances between the left channels and between the right channels of the recordings. Distances were calculated for the recordings of each of the categories separately, as well as for the categories combined (*All*). In general, the versions denoted SBD are clustered much closer together, whereas we get a wider variety among AUD recordings. MAT recordings vary less than AUD but more than SBD. Some of the fluctuations, such as the higher SBD distances in the first two examples in the table are likely a consequence of incorrect annotation of the data.

6 Conclusion and Future Work

In this paper we presented a system that automatically aligns and clusters live music recordings based on spectral audio features and editorial metadata. The procedure presented has been developed in the context of a project aiming at demonstrating how Semantic

Concert date	Song	All	AUD	SBD	MAT
1982-10-10	Playing In The Band	0.388543	0.410841	0.277323	0.375578
1982-10-10	Throwing Stones	0.367355	0.392155	0.23965	0.276406
1983-10-15	Playing In The Band	0.230666	0.238601	0.010321	n/a
1983-10-15	Throwing Stones	0.224912	0.225901	0.010876	n/a
1985-07-01	Good Lovin'	0.183972	0.175587	0.033591	n/a
1985-07-01	Playing In The Band	0.18329	0.176066	0.033946	n/a
1987-09-18	Sugaree	0.181573	0.163099	0.06254	0.215186
1987-09-18	Walking Blues	0.135	0.177194	0.050736	0.079859
1989-10-09	Playing In The Band	0.13585	0.135958	0.028306	0.076555
1989-10-09	Throwing Stones	0.167371	0.170313	0.030709	0.075483
1990-03-28	Good Lovin'	0.167065	0.161243	0.018454	0.122656
1990-03-28	Hey Pocky Way	0.166781	0.13595	0.014054	0.143182
1990-03-29	The Wheel	0.375285	0.414431	0.009073	0.505686
1990-03-29	Throwing Stones	0.346342	0.347416	0.024085	0.492202
1990-07-14	Crazy Fingers	0.21509	0.242463	0.155493	0.008562
1990-07-14	Throwing Stones	0.134404	0.15944	0.007485	0.013143
average		0.225219	0.232916	0.062915	0.198708

Table 2: Average distances for different stereo recordings of one song performance.

Audio and Linked Data technologies can produce an improved user experience for browsing and exploring music collections online. It will be made available in a prototypical Web application that links the large number of concert recordings by the Grateful Dead available in the Internet Archive with audio analysis data and retrieves additional information and artefacts (e.g. band lineup, photos, scans of tickets and posters, reviews) from existing Web sources, to explore and visualise the collection.

We demonstrated how the system discussed in this paper can help us understand the material and evaluate it against the information given by the online community. Potentially, such procedures can not only be used to complement incomplete data or correct annotation errors, but also to discover previously unknown relationships between audio files. Future work includes developing similar procedures for other musical material, such as versions of the same song played at different concerts, and further research into algorithms for the alignment of different audio sources.

Acknowledgments

This paper has been supported by EPSRC Grant EP/L019981/1, Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption.

References

[1] Benson, M., *Why the Grateful Dead Matter*, ForeEdge Press, 2016.

[2] Meriwether, N., "Documenting the Dead," *Online: http://www.dead.net/documenting-the-dead*, 2015.

[3] Bell, M., "Guide to Cassette Decks and Tape Trading," *Online: https://www.cs.cmu.edu/ ~./gdead/taping-guide/taping-guide.txt*, 1995.

[4] Bechhofer, S., Page, K., and De Roure, D., "Hello Cleveland! Linked Data Publication of Live Music Archives," in *Proceedings of WIAMIS, 14th International Workshop on Image and Audio Analysis for Multimedia Interactive Services*, 2013.

[5] Bechhofer, S., S.Dixon, Fazekas, G., Wilmering, T., and Page, K., "Computational Analysis of the Live Music Archive," *Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR 2014)*, 2014.

[6] Wilmering, T., Fazekas, G., Dixon, S., Bechhofer, S., and Page, K., "Automating Annotation of Media with Linked Data Workflows," in *3rd International Workshop on Linked Media (LiME 2015) co-located with the WWW'15 conference, 18-22 May, Florence, Italy.*, 2015.

[7] Bartsch, M. A. and Wakefield, G. H., “To Catch a Chorus: Using Chroma-based Representations for Audio Thumbnailing,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 15–18, New Paltz, NY, USA, 2001.

[8] Davis, S. and Mermelstein, P., “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE transactions on acoustics, speech, and signal processing*, 28(4), pp. 357–366, 1980.

[9] Foote, J. T., “Content-Based Retrieval of Music and Audio,” in C.-C. J. Kuo, S.-F. Chang, and V. N. Gudivada, editors, *Multimedia Storage and Archiving Systems II*, 1997.

[10] Logan, B. and Salomon, A., “A Music Similarity Function Based in Signal Analysis,” *IEEE International Conference on Multimedia and Expo (ICMC)*, 2001.

[11] Mitrović, D., Zeppelzauer, M., and Breiteneder, C., “Features for Content-Based Audio Retrieval,” *Advances in Computers*, 78, 2010.

[12] Dixon, S. and Widmer, G., “MATCH: A Music Alignment Tool Chest.” in *ISMIR*, pp. 492–497, 2005.

[13] Borg, I. and Groenen, P. J., *Modern multidimensional scaling: Theory and applications*, Springer Science & Business Media, 2005.

[14] Thalmann, F., Perez Carillo, A., Fazekas, G., Wiggins, G. A., and Sandler, M., “The Mobile Audio Ontology: Experiencing Dynamic Music Objects on Mobile Devices,” in *Tenth IEEE International Conference on Semantic Computing*, Laguna Hills, CA, 2016.

[15] Harris, M., Smaill, A., and Wiggins, G., “Representing Music Symbolically,” in *Proceedings of the IX Colloquio di Informatica Musicale*, Venice, 1991.

[16] Thalmann, F., Perez Carillo, A., Fazekas, G., Wiggins, G. A., and Sandler, M., “The Semantic Music Player: A Smart Mobile Player Based on Ontological Structures and Analytical Feature Metadata,” in *Web Audio Conference WAC-2016*, Atlanta, GA, 2016.