GUITAR TAB MINING, ANALYSIS AND RANKING

Robert Macrae
Centre for Digital Music
Queen Mary University of London
robert.macrae@eecs.qmul.ac.uk

Simon Dixon
Centre for Digital Music
Queen Mary University of London
simon.dixon@eecs.qmul.ac.uk

ABSTRACT

With over 4.5 million tablatures and chord sequences (collectively known as tabs), the web holds vast quantities of hand annotated scores in non-standardised text files. These scores are typically error-prone and incomplete, and tab collections contain many duplicates, making retrieval of high quality tabs difficult. Despite this, tabs are by far the most popular means of sharing musical instructions on the internet. We have developed tools that use text analysis and alignment for the automatic retrieval, interpretation and analysis of such tabs in order to filter and estimate the most accurate tabs from the multitude available. We show that the standard means of ranking tabs, such as search engine ranks or user ratings, have little correlation with the accuracy of a tab and that a better ranking method is to use features such as the concurrency between tabs of the same song. We also compare the quality of top-ranked tabs with state-of-the-art chord transcription output and find that the latter provides a more reliable source of chord symbols with an accuracy rate 10% higher than the ranked hand annotations.

1. INTRODUCTION

There are a number of digital music notation formats, such as Music XML, the MIDI file format, and various formats for images of scanned sheet music. However it is tabs, which are plain text files containing tablature and/or chord symbols and lyrics, that have become the most commonly used music notation format on the internet. A comparison of the most popular MIDI, sheet music and tab websites' unique visitors per month can be seen in Table 1. The popularity of tabs is due to a simple, intuitive approach to the instructions that requires no formal training to understand nor specific software to read or write. Added to this is the fact that tabs are commonly free to use and the amount of data needed to transfer the text instructions is almost negligible. However, due to the lack of standardisation there are many variations in how they are structured, making machine parsing of tabs difficult. Also, since even beginners can use the format to annotate music, many of the tabs found are of poor quality, suffering from errors and incompleteness. A further problem is that multiple tabs exist for many songs, making it difficult for the user to locate the most accurate and complete instance among the alternatives. These difficulties motivate the current work.

File type	Most popular site	Visitors
tabs	ultimate-guitar.com	2,541,482
sheet music	8notes.com	470,010
MIDI	freemidi.org	17,437

Table 1. Unique visitors per month to music score websites from http://siteanalytics.compete.com

We address these problems by developing a parser for guitar tabs and using music information retrieval methods to analyse and compare the tabs. We propose several features and evaluate their effectiveness as predictors of tab accuracy, in order to improve the quality of tab retrieval. Overall, we aim to evaluate the viability of data-mining a noisy source of metadata from the internet, and we compare our results with those obtained by content-based analysis of audio for determining the chord sequence for a given song.

Despite the popularity of tabs on Usenet groups such as alt.guitar.tab in the 1990's and more recently on web sites such as ultimate-guitar.com, little attention has been given to this source of data by the music information retrieval community. In recent work McVicar and De Bie [8, 9] showed how chord sequences from guitar tabs, synchronised with the music, can help improve machine learning methods for chord recognition. Audio and video analysis were used in [3] to find the simplest tablature transcription of chords and a guitar tablature score follower was demonstrated in [5] that used score following to display tabs on small screens.

2. THE BEATLES DATA

In this work we focus on The Beatles due to the availability of ample annotated data and guitar tabs for this band.

2.1 Ground Truth Chord Sequences

The ground truth chord sequence annotations for The Beatles used in this work come from transcriptions by Chris Harte [2]. This data includes chord sequences for 180 tracks from 12 Beatles' studio albums, which is the set that we focus on in our evaluation.

2.2 Ground Truth Structure Segmentation

These are structural segmentations consisting of start time, end time and segment label for the same 180 The Beatles tracks [6]. The labels are words such as *verse, refrain, bridge, intro, outro,* and *silence*, which are often qualified with details, e.g. *verse_a, verse_b,* and *verse_(guitar_solo).*

2.3 Web-Mining

We used two search engines to locate guitar tabs. The first, 911tabs.com, is a guitar tab search engine with over 4.5 million tabs indexed. We wrote a web crawler that retrieved all the correctly labelled Beatles tabs from 911tabs.com corresponding to the 180 tracks in our test set. For the second search engine, Google, we found a combination of search terms ('guitar', 'tab' and some filters for unwanted content such as '-video') that, when combined with the artist and track name, gave a high ratio of tabs in the results. After the first 100 results for each tab search, the number of tabs returned was low, so we focused on the top 100 results for each song. In total we found 24746 tabs relating to the 180 Beatles tracks in our ground truth data. Additionally, we mined the web for an initial chord dictionary of 264 common chords from chordie.com and guitarsite.com.

3. TAB PARSING

We see decoding tabs as an example of noisy text analytics, which are often applied to determine meaning from web-mined resources such as online chat, forums, blogs and wikis. To interpret the noisy semi-structured tab data, we implemented a large set of simple heuristics to handle the many varied tab writing styles that exist. The following steps are a brief outline of the stages involved in parsing tabs.

- Interpret any HyperText Markup Language (HTML) specific tags. For instance,   and
 tags are changed to spaces and new lines, respectively.
- Analyse each line to determine what (if any) type of tab line it is. For example the line could contain a 'structural marker', 'chord line', 'chord and lyrics line', 'tablature line', etc. Non-tab-specific text is discarded.
- For each tab line, decode the tab elements accordingly. As such, chords will be extracted from any 'chord line' or 'chord and lyrics line', notes will be extracted from a 'tablature lines', new chords will be added to the tabs chord dictionary from any 'chord definition line'.
- Reorganise the tab sections into an organised tab according to given structural information. Any indicators of repetitions will be expanded so that 'x2' will result in the current section being duplicated.

We developed our heuristics for parsing guitar tabs on a set of 20 tabs for which we manually annotated ground truth. The chord retrieval from these tabs, as an example, extracts 806 out of the 807 chords correctly. Figures 1 - 2 demonstrates two different samples of tab formats along with the chords extracted by our parsing tool in each case.

```
[Intro]
Riff1
e------0-|-3---3---5---5-|-10-------------------|-------8---------|
B---3--1-|-3---3---7---7-|-12----12-0--0--12-0--|-------10--10--7-|
G--------|-4---4---7---7-|-12----12-12-12-12-12-|-9----9----------|
D--------|---------------|----------12-12-12-12-|-10--------------|
A--------|---------------|----------------------|(10)-------------|
E--------|---------------|----------------------|-----------------|

Riff2
e--3--3--3--3--|--0--0------(0)-|
B--3--3--3--3--|--3--3--(3)-----|
G--0--0--2--2--|--0--0----------|
D--0--0--0--0--|--2--2---0---2--|
A--2--2--x--x--|--2--2----------|
E--3--3--2--2--|--0--0----------|
   G     D/F#     Em

G        D/F#    Em
Love     love    love
G        D/F#    Em
Love     love    love
D7/A     G       D7/F#    D7/E
Love     love    love
D    C    Riff3
```

Figure 1. Tab Sample 1. Chords Extracted: G D/F# Em G D/F# Em G D/F# Em D7/A G D7/F# D7/E D C

```
A taste of [Am]honey, [C]tasting much [G]sweeter than [Am]wine

I [Am]dream of [C]your first [G7]kiss and [D]then
I [Am]feel a[C]part, my [G7]lips are [D]gett'n
A taste of [Am]honey, [C]tasting much [G]sweeter than [Am]wine

{Chorus:}
I [A]will re[C]turn, yes [D]I will re[Em]turn
I'll come [F]back for the [G]honey and [Am]you.
```

Figure 2. Tab Sample 2. Chords Extracted: Am C G Am Am C G7 D Am C G7 D Am C G Am A C D Em F G Am

4. EVALUATION

In this section we evaluate the precision of the tabs themselves and then compare various means of ranking the tabs. In order to do this, we first describe the features used for measuring and predicting the tabs' accuracies. We also explain existing ranking methods, such as 911.com's user rating and Google's page rank. We then use correlation to determine the suitability of using these features as ranking methods. Also, for each feature, we compare the selected chord sequences with the output of a state-of-the-art automatic chord detection system.

	C#	C#6	Db	Fm7	C/B	A11	Dm/C#
C#	0.0	0.25	0.0	0.5	1.0	0.8	0.75

Table 2. Chord Similarity (CS) cost examples.

4.1 Features

4.1.1 Chord Similarity (CS)

In order to measure two chords' similarity, we use the Levenshtein Distance (LD) [4] of the alphabetically ordered notes in the chord, as interpreted from the chord definitions. The LD uses dynamic programming to find a path $P(U,V) = (p_1, p_2, ..., p_W)$ through a matrix of costs between sequences $U = (u_1, u_2, ..., u_M)$ and $V = (v_1, v_2, ..., v_N)$. This cost matrix is described as $d_{U,V}(m,n)$ where $m \in [1:M]$ and $n \in [1:N]$ where each $p_k = (m_k, n_k)$. LD uses a cost of 0 for matches and 1 for any insertion, deletion or alteration. The maximum cost is the length of the longest sequence. We normalise and invert this cost to give a similarity value from 0 to 1, between two chords (note sequences), U and V.

$$\text{CS(U,V)} = \left(1 - \frac{\text{LD}(U,V)}{\max(M,N)}\right) \quad (1)$$

Due to how tab parser interprets chord definitions, this cost function treats any enharmonic chords or notes equally. Examples of this cost function (CS) can be seen in Table 2.

4.1.2 Chord Sequence Similarity (CSS)

The Chord Sequence Similarity is a measure of how similar two tab chord sequences, T_1 and T_2 are. For this method we use DTW, a generalisation of LD, which has been used for synchronisation in applications such as score following [1]. Unlike the binary comparison in LD, DTW can use a more detailed cost function such as the inner product of the pair of feature vectors, which returns a value between 0 and 1 for each pair of feature vectors. In our case the DTW uses the CS cost function to compare chords. The overall similarity cost is given by the sum of the individual chord match costs along the DTW path P and the maximum cost is the length of the longest sequence. We normalise and invert this similarity cost and express it as a percentage:

$$\text{CSS}(T_1,T_2) = \left(1 - \frac{\text{DTW}(T_1, T_2, CS)}{\max(|T_1|, |T_2|)}\right) \times 100 \quad (2)$$

Examples of the CSS can be seen in Table 5.

4.1.3 Chord Accuracy (CA)

The Chord Accuracy measures the similarity of the overall sequence of chords T in a tab to the chord sequence G in the ground truth data for the song. Transpositions are not considered in this factor.

$$\text{CA}(T,G) = \text{CSS}(T,G) \quad (3)$$

4.1.4 Segment Chord Accuracy (SCA)

Many tabs have incomplete chord sequences, and rely on the user to piece together the complete tab based on cues, intuition and knowledge of the song. A more flexible accuracy measurement, the Segment Chord Accuracy, finds the accuracy of each segment in the tab independently. For each structural segment of a song, as defined in our structural ground truth data, the SCA takes the closest matching sub-sequence from the tab's overall chord sequence. In addition, chord sub-sequences which match to more than one segment may be reused and transpositions of the data are allowed in the SCA measurement. The pseudo-code for the SCA is shown in Algorithm 1.

```
Input: Segmentation S = {s1, s2, ..., sl}, Ground Truth
       Chords G, Tab Chords T
Output: Segment Chord Accuracy SCA
SCA = length(G);
for Transposition Tr = 0 to 11 do
    TranspositionCost = 0;
    for i = 1 to l do
        SegCost = length(si);
        for start = 0 to length(T) do
            for len = 1 to length(T) − start do
                T′ = subsequence(T,start,len)
                if CSS(si, T′) < SegCost then
                    SegCost = CSS(si, transpose(T′, Tr));
                end
            end
        end
        TranspositionCost += SegCost ;
    end
    if TranspositionCost < SCA then
        SCA = TranspositionCost;
    end
end
return SCA;
```

Algorithm 1: Segment Chord Accuracy

4.1.5 Chords Concurrence (CC)

To determine how well tabs of a song agree with each other, we define the Chords Concurrence as the average of the similarities between a tab's chord sequence T_k and the chord sequences $T_i (i \neq k)$ of all the other tabs of the same song.

$$\text{CC}(T_k) = \sum_{i=1, i \neq k}^{n} \text{CSS}(T_k, T_i)/(n-1) \quad (4)$$

4.1.6 Structure Similarity (SS)

In order to calculate Structure Similarity we first normalise the labelling of structural segments, so that a musical structure such as (Intro, Verse, Chorus, Verse,...) is represented by the sequence of characters (A, B, C, B, ...). We then use the LD, normalised and inverted to provide a percentage:

$$\text{SS}(T_1, T_2) = \left(1 - \frac{\text{LD}(T_1, T_2)}{\max(T_1, T_2)}\right) \times 100 \quad (5)$$

Note that we only compute Structure Similarity where the structure is explicitly given in the tab.

4.1.7 Structure Accuracy (SA)

The Structure Accuracy is a measure of how similar the structural sequence T of a tab is to the structural sequence G of the ground truth data.

$$\mathrm{SA}(T) = \mathrm{SS}(T, G) \tag{6}$$

4.1.8 Structure Concurrence (SC)

The Structure Concurrence is the average of the similarities between a tab's structural sequence T_k and the structural sequences T_i of all the other tabs of the same song.

$$\mathrm{SC}(T_k) = \sum_{i=1, i\neq k}^{n} \mathrm{SS}(T_k, T_i)/(n-1) \tag{7}$$

4.1.9 911 Rating

The 911 Rating is the average user rating assigned to the tab at www.911tabs.com from 1 (bad) to 5 (good). The number of votes that went into this average rating is not provided by the tab site. 1246 tabs with chords had 911 Ratings.

4.1.10 Google Rank

The tab's Google Rank corresponds to where the URL of the tab is found in the ordered list of Google's ranked search results [10]. Values range from 1 (best) to 100 (worst known). 5619 tabs found had Google Ranks associated with them, 1931 of which had chord sequences.

4.1.11 Date Modified

If posted tabs are edited and reposted, it might be the case that more recent tabs are more accurate on average than earlier tabs. A tab's Date Modified is the last modified value of the HTML file on the tab server, expressed as the number of milliseconds since 00:00:00 January 1, 1970 GMT. 2022 of the tabs with chord sequences had an associated last date modified.

4.2 Guitar Tab Statistics

Of the 24746 tabs found with our web-mining tool, 7547 had recognisable chord content and 4643 had structure explicitly defined, with at least 3 chords/sections. The average tab Chord Accuracy (CA) for tabs, tabs that were duplicates and non duplicates is 61.8%, 63.4%, and 58.3% respectively. A similar pattern was observed in the Structure Accuracy (SA) of 50.0%, 50.3%, and 49.1%, suggesting that more accurate tabs are more likely to be copied. The accuracy difference is however small, and the Pearson-rank correlation shows a very weak correlation between accuracy and whether a tab is duplicated (0.12 for CA and 0.03 for SA).

Filter Method	Pearson-rank correlation		Samples
	CA	SCA	
Chords Concurrence	0.54	0.51	7547
911 Rating	0.07	0.06	1161
Google Rank	-0.07	-0.08	1935
Date Modified	0.03	0.01	2022

Table 3. Correlations between various features and the Chord Accuracy (CA) and Segment Chord Accuracy (SCA).

Filter Method	Pearson-rank correlation	Samples
	SA	
Structure Concurrence	0.19	4643
911 Rating	0.02	620
Google Rank	-0.07	1197
Date Modified	0.06	1337

Table 4. Number of samples and correlation values between various features and the Structure Accuracy (SA).

4.3 Tab Ranking Systems

The Pearson-rank correlation is an indication of how effective a ranking system is. For example, if there is a high and statistically relevant correlation between a tab's score in its 911 Rating and its CA, we can deduce the 911 Rating favours accurate tabs. Table 3 shows the correlations found between the tabs' CA, SCA and 4 relevant features discussed above. Similarly, we give the correlations with the Structure Accuracy in Table 4.

Two of the correlations from Table 3 can be seen in the scatter plots in Figures 3 and 4. Each point represents the CA (vertical coordinate) plotted against a feature value (horizontal coordinate) for a single tab. The features used are Google Rank (Figure 3) and CC (Figure 4). A negative correlation in Figure 3, shows that tabs with higher Google Ranks (lower numbers) are more accurate. A stronger trend can be seen in Figure 4, where the tabs with a higher Chord Concurrence have a higher Chord Accuracy. Figure 5 shows the correlation between Structure Accuracy and Structure Concurrence from Table 4. Again, there is a clear trend between concurrence and accuracy.

For the sample sizes provided; the required absolute value for statistical significance is less than 0.1. Surprisingly, the rating given by users at 911tabs.com, the date the tab was made and the Google Rank had no statistically significant correlation with the accuracy of the tab. The strongest correlation was provided by the Concurrence methods that had a Pearson-rank correlation of 0.54 for CA, 0.51 for SCA, and 0.33 for SA.

These results show it is possible to improve the ranking of tabs by search engines based on analysing the contents of tabs in relation to other tabs of the same track.

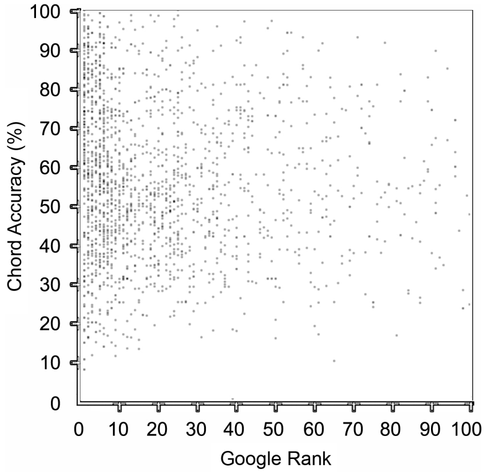

Figure 3. The scatter plot relates the Chord Accuracy to the Google Rank. Note that lower numbers correspond to higher rank. The weak negative trend between the Google Rank and accuracy of the tabs is not significant.

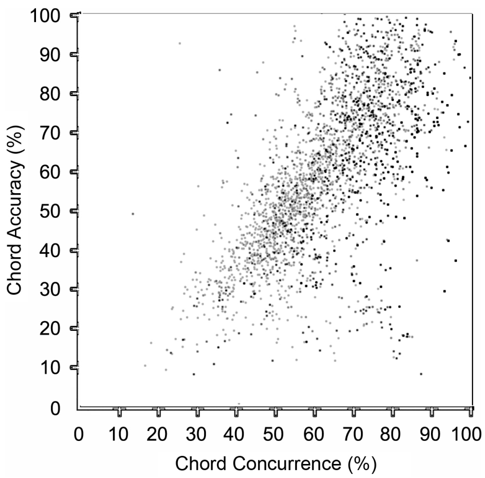

Figure 4. This scatter plot shows a strong trend relating Chord Accuracy and Chord Concurrence.

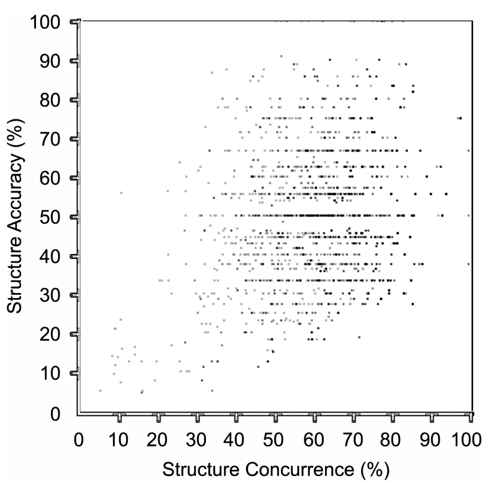

Figure 5. This graph shows the trend between the Structure Accuracy and the Structure Concurrence.

4.4 Automatic Chord Transcription

Our final experiment was to compare the results with automatic chord detection methods. Both methods satisfy the same information need: finding the chords to a given song. We selected the top ranking tab for each feature and compared the accuracy of its chord sequence with the output of a state-of-the-art automatic chord detection system [7].

In Table 5 there is an example of the chord sequences produced by the automatic chord recognition system, those selected by our features, and the ground truth annotations for The Beatles' *Don't Pass Me By*. The chord accuracies are also given. Table 6 shows the average accuracy of the methods. There is a clear superiority in the automatic detection algorithm which is over 10% more accurate, on average, than the tabs selected by any of our features. Of the features, the Chord Concurrence is the most successful feature for selecting the tab to use. Additionally, we can improve results by selecting only tabs with a high Chord Concurrence value. For example, those with 90% CC or more have an average Chord Accuracy of 79.9%. However, only 24 out of 7547 tabs have such a high Chord Concurrence.

4.5 Dependance on Sample Size

A potential weakness of the Concurrence methods could be in being dependent on the number of tabs available for a particular song. To see if this would effect performance, we calculated the correlation between N (the number of tabs for a particular song) and C (correlation between CA and CC) for each of our 180 tracks. The result, 0.039, is not statistically significant for the sample size, suggesting that Chord Concurrence is a relevant indicator of Chord Accuracy regardless of the number of tabs on which it is based.

5. DISCUSSION AND FUTURE WORK

Using tab concurrence, we are able to order tab search results so that the more accurate tabs are given preference, thereby improving the tab search experience. If ranking tabs based on one feature leads to a clear improvement over current ranking systems, it is possible that greater improvements can be made by selecting tabs using more sophisticated combinations of features. Whilst we have limited ourselves to analysing just the guitar tabs themselves, we see possible synergies in combining this work with other projects based on web-mining multimodal music metadata [11], content-based analysis [7], and other scores.

The usefulness of Chord Concurrence is not surprising, as errors are less likely to be replicated in independently produced tabs, than the correct chords. However, the automatic transcription result shows that a machine listening method performs better than the average human annotator, and this result holds even when features are used to select

Source	CA	Chord Sequence
Ground Truth	-	C C F G F C C F G F C C F C G F C C F G F C C F C G F C C/5 C C F C G F C F G Csus4 C C
Auto (Mauch)	90.4%	C C F G F C/5 F G F C Cmin7 F C G F C F G F C/5 F C G F C G Gmaj6 G7 C Cmin C F C G F C F G C C/5 F
Chord Concurrence	89.4%	C F G F C C F G F C F C G G F C F G F C F C G F C
911 Rating	82.9%	C F G F C C F G F C C F C G F C F G F C
Google Rank	63.4%	G C D C G G C D C G G C G D C G G C D C G G C G D C G G C G D C G
Date Modified	63.4%	G C D C G G C D C G G C G D C G G C D C G G C G D C G G C G D C G

Table 5. Example chord sequences retrieved by the various chord detection methods for the song *Don't Pass Me By* showing the Chord Accuracy (CA) of these sequences.

Detection Method	Chord Accuracy
Auto (Mauch)	79.3%
Chord Concurrence	68.8%
911 Rating	66.9%
Google Rank	65.6%
Date Modified	62.3%
Randomly Selected	61.8%

Table 6. The average Chord Accuracy of the chord sequences, over 180 Beatles tracks, that were provided by the top-ranked tabs and the chord detection methods. The final row shows the average as if the tab was randomly selected.

better-than-average tabs. This raises an interesting question about ground truth: To what extent can human annotations from unknown sources be used as ground truth in MIR?

In future work we plan to improve on the ranking techniques demonstrated here for the purposes of recommendation, synchronisation, tab generation and score following. This work has shown that the concurrence of tabs indicates their accuracy, therefore we hypothesise that concurrency in subsequences and tablature notation will follow this rule. The prevalence of tabs and the tools described here present many interesting avenues of research, including artist similarity, the use of chord idioms and influences across genres. Our experiments show, with others [8,9], that tabs are a useful source of data for research in MIR.

6. ACKNOWLEDGMENTS

The authors thank Matthias Mauch for the structural annotations, automatic chord transcription, and review of this work. We also thank Chris Harte for the chord annotations. This work was funded by an EPSRC DTA studentship.

7. REFERENCES

[1] R. B. Dannenberg. An on-line algorithm for real-time accompaniment. In *International Computer Music Conference*, pages 193–198, 1984.

[2] C. Harte, M. Sandler, and S. Abdallah. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proc. 6th International Conference on Music Information Retrieval (ISMIR)*, pages 66–71, 2005.

[3] Alex Hrybyk and Youngmoo E Kim. Combined audio and video analysis for guitar chord identification. In *Proc. 11th International Society on Music Information Retrieval (ISMIR)*, pages 159–164, 2010.

[4] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics*, 10(8):707–710, February 1966.

[5] R. Macrae and S. Dixon. A guitar tablature score follower. In *IEEE International Conference on Multimedia & Expo (ICME)*, pages 725–726, 2010.

[6] M. Mauch, C. Cannam, M. Davies, S. Dixon, C. Harte, S. Kolozali, D. Tidhar, and M. Sandler. OMRAS2 metadata project 2009. In *Late-breaking session at the 10th International Society on Music Information Retrieval (ISMIR)*, 2009.

[7] M. Mauch and S. Dixon. Simultaneous estimation of chords and musical context from audio. *IEEE Trans. Audio, Speech and Lang. Proc.*, 18:1280–1289, 2010.

[8] Matt McVicar, Yizhao Ni, Raul Santos-Rodriguez, and Tijl De Bie. Leveraging noisy online databases for use in chord recognition. In *Proc. 12th International Society on Music Information Retrieval (ISMIR)*, 2011.

[9] Matt McVicar, Yizhao Ni, Raul Santos-Rodriguez, and Tijl De Bie. Using online chord databases to enhance chord recognition. *Journal of New Music Research*, 40(2):139–152, 2011.

[10] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[11] M. Schedl, G. Widmer, P. Knees, and T. Pohle. A music information system automatically generated via web content mining techniques. *Information Processing & Management*, 47(3):426–439, 2011.