COMPARING CQT AND REASSIGNMENT BASED CHROMA FEATURES FOR TEMPLATE-BASED AUTOMATIC CHORD RECOGNITION

Ken O'Hanlon, Mark B. Sandler

Centre for Digital Music
Queen Mary University of London

ABSTRACT

Automatic Chord Recognition (ACR) seeks to extract chords from musical signals. Recently, deep neural network (DNN) approaches have become popular for this task, being employed for feature extraction and sequence modelling. Traditionally, the most important steps in ACR were extraction of chroma features which estimate the energy in each pitch class, and pattern matching using templates or learning-based approaches. In this paper we reconsider chroma features with template matching, employing spectral reassignment chroma with synthetic spectral templates, and find experimental results comparable to those of a recent DNN-based chroma extractor.

Index Terms— chord recognition, spectral reassignment, chroma, CRP feature

1. INTRODUCTION

Chords are a fundamental construct in much western music and their automatic recognition is a popular task in music processing. Traditionally, ACR was performed by first calculating chroma features [1], in which each dimension relates an estimate of activity in a given pitch class. A chromagram is a matrix with a chroma feature associated to a time instant in each column. Chord labelling is then performed by comparing the chroma features to models for different chords. Chord models may be templates, or machine learning based. Chroma features are noisy due to the presence of transients and other non-chord related signal elements. Hence, ACR systems invariably incorporate temporal continuity [2] [3] Other music features, including key [4], bass [5], beats [3] [6], have been explored to counter the noise in chroma. Even so, the importance of features in ACR was asserted by Cho and Bello [7] who compared different chroma and found ordered performance to be almost invariant to various post-processing steps. Currently, deep neural networks (DNN) are popular in ACR, for feature extraction [8] [9] and temporal consideration [10] [9], although the delineation of ACR subtasks may be blurred in DNNs [11].

This research was funded by the EPSRC Program Grant EP/L019981/1. Mark Sandler acknowledges a Wolfson Merit Fellowship.

A variety of chroma features have been proposed for ACR and other tasks. Typically a pitch feature is first derived from which a basic chroma feature is calculated by summing elements of the same pitch class. The pitch feature is usually based on a Constant Q-transform (CQT), a spectrogram with a log-frequency scale with an equal number of frequency bins per octave. Alternative features to this basic chroma are easily derived through transformation of the pitch feature. Well-known transforms for chroma include log-compression which de-emphasises large energy peaks and spectral weighting which places more weight on the central elements of the pitch scale, thereby de-emphasising the effect of higher overtones. A chroma feature combining these two transforms applied to a CQT-based pitch feature was seen to be optimal in [7]. Later, we found the Chroma Reduced Pitch (CRP) feature [12] which high-pass filters the log-compressed pitch feature, to improve on the log chroma feature [13]. While chroma features have been often compared [7] [14], the effect of different pitch features has not been explored so thoroughly. One alternative spectral representation to CQT is spectral reassignment, employed in [15].

Different chord modelling techiques have been explored for ACR. The original [1], and simplest approach is to use binary chord templates, in which expected active pitch classes are set to one. Data-driven chord models later became widely used, particularly multivariate Gaussian and Gaussian mixture models (GMM), [5] [16] [7]. Other learning-based approaches include SVMs [17] [18]. A synthetic template was employed for chord modelling in [19] [6]. In this case a chord template was formed by summing model note spectra [20] synthesised according to a number of overtones with a fixed roll-off. Such synthetic templates were only applied to basic chroma and were found to have little effect [19] relative to the binary template and have been rarely employed since.

In this paper we reconsider synthetic template-based ACR for log and CRP features. We derive a CQT through reassignment and compare to a standard CQT for various chroma. In the next section we detail the approach taken, including reassignment, chroma, and synthetic chord models. We then provide experimental results for chroma features using both CQTs, and report results similar to the deep chroma extractor [8]. Finally we conclude, with pointers to future work.

2. PROPOSED APPROACH

2.1. Reassigned CQT and pitchgram

Spectral reassignment, which assigns the energy in a point on the spectrogram $S(\omega,\tau)$ calculated with a window, w, to a point $S(\hat{\omega},\hat{\tau})$ is first performed, similar to [15] using the reassignment operators of [21], in which

$$\hat{\omega} = \omega + \frac{\partial\phi(\omega,\tau)}{d\tau} = \omega + \Im\left(\frac{S_D(\omega,\tau)\times S^*(\omega,\tau)}{|S(\omega,\tau)|^2}\right) \quad (1)$$

is proposed for frequency assignment, where $\phi(\omega,\tau)$ is the phase at the spectrogram point $S(\omega,\tau)$, S_D is a spectrogram calculated using a window, $w_D = \frac{dw(t)}{dt}$, and S^* is the complex conjugate of the spectrogram. Likewise,

$$\hat{\tau} = \tau - \frac{\partial\phi(\omega,\tau)}{d\omega} = \tau - \Re\left(\frac{S_T(\omega,\tau)\times S^*(\omega,\tau)}{|S(\omega,\tau)|^2}\right) \quad (2)$$

performs temporal reassignment, where S_T is a spectrogram calculated using a window, $w_T = tw(t)$, The mixed second derivative [22] is also calculated

$$\frac{\partial^2\phi(\omega,\tau)}{\partial\tau\partial\omega} = -\frac{\partial\hat{\tau}(\omega,\tau)}{\partial\tau} = \frac{\partial\hat{\omega}(\omega,\tau)}{\partial\omega} - 1 = \Re\left(\frac{S_{TD}(\omega,\tau)S^*(\omega,\tau)}{|S(\omega,\tau)|^2}\right) - \Re\left(\frac{S_T(\omega,\tau)S_D(\omega,\tau)}{S^2(\omega,\tau)}\right) \quad (3)$$

where S_{TD} is a spectrogram calculated using a window $w_{DT} = w(t)\times\frac{\partial w(t)}{\partial t}$. This is used to filter out transient elements by thresholding [15]

$$\Lambda = \left\{(\omega,\tau)\left|\left|\frac{\partial^2\phi(\omega,\tau)}{\partial\tau\partial\omega}+1\right|<0.4\right.\right\} \quad (4)$$

as it is shown that $\frac{\partial^2\phi(\omega,\tau)}{\partial\tau\partial\omega} = -1$ for sinusoids and $\frac{\partial^2\phi(\omega,\tau)}{\partial\tau\partial\omega} = 0$ for clicks [22].

After calculating the reassignment operators (1) (2) (3), we can derive a reassigned CQT (RA-CQT). This is designed in order to be as similar to the CQT employed in [7], and a time-frequency grid with 36 bins per octave, or 3 bins per semitone, and with $92ms$ temporal resolution is created. The frequency points on the grid are generated according to an estimated tuning, using a pitch error histogram-based method [20] using the reassigned frequencies of high energy points in the spectrogram. The same tuning estimations can also be used for the standard CQT filterbank. For each point in the set Λ, the energy is assigned into the grid according to the reassigned frequency and time. A pitch representation $\mathbf{P}\in\mathbb{R}^{120\times N}$ is then calculated with each row representing a pitch. Rows 21 to 108 are populated by the pitches corresponding to their MIDI number, covering the scale of a piano, while other rows are zero, as used for CRP [12]. As the RA-CQT contains 36 bins per octave, a gaussian weighting is applied, across the three RA-CQT bins representing one point in $\mathbf{P}$ in order to downweight the sidelobes, similar to the CQT used in [7]. A large temporal overlap in the spectrogram produces a smoother RA-CQT in the presence of noise and tranisents.

2.2. Chroma features

Once the basic pitchgram is computed, the transforms can be applied such as log compression

$$p^L_{m,n} = log(1+\alpha p_{m,n}) \quad (5)$$

where $\alpha = 1000/\max(\mathbf{p}_n)$ and $p_{m,n}$ is the element in the mth row in $\mathbf{p}_n$, the nth column of P . Spectral weighting such as in [7] uses a Gaussian centred on $m = 60$

$$p^W_{m,n} = p_{m,n}\times e^{-\frac{(m-60)^2}{450}} \quad (6)$$

A reduced pitch feature, such as CRP is formed

$$\mathbf{p}^R_n = \mathcal{H}(\mathbf{p}_n) \quad (7)$$

where $\mathcal{H}$ is a high-pass filter, effected through removing a set of low-frequency coefficients from a DCT of dimension 120 [12]. Different sets can be used, typically removed from each other by 20 dimensions e.g. the CRP(15), CRP(35) and CRP(55) remove the lowest $15, 35$, and 55 coefficients, respectively. A jump effect is seen at this distance of 20 which is also the distance between cosine elements that repeat on a per-octave basis in the specified DCT e.g. DCT(41) repeats twice per octave, while DCT (81) repeats 4 times per octave. While CRP(55) is considered favourable for audio similarity in [12], we found [13] that CRP(35) was superior for ACR.

The various transforms (5) (6) (7) can be performed in combination with each other in one feature. Here, the log compression is performed first (5) followed by the filtering (7) with the spectral weighting (6) always performed last. Finally the chroma is estimated by addition of the coefficients of the different pitch classes across all octaves

$$c_{k,n} = \sum_{o=2}^{O+1} p_{12\times o-4+k,n} \quad (8)$$

where $O = 7$ is number of octaves considered here.

2.3. Templates

Binary templates are formed by creating a chroma feature with pitch classes expected active in a given chord set to one and all other dimensions set to zero. For example an A major chord template is specified as

$$t = [1,0,0,0,1,0,0,1,0,0,0,0]^T$$

which can be cycled to give all other major templates. Minor chord templates are formed similarly. A template dictionary $\mathbf{T}\in\mathbb{R}^{12\times l}$, where l is the number of chords considered is formed by placing a labelled chord template in each column.

A dictionary of synthetic templates is defined by two parameters; the number of harmonics, h, and a roll-off parameter, r. The amplitude of the vth harmonic in a note is then defined as

$$a^v = r^{v-1}.$$

While previous approaches [19] [6] simply add ideal expected spectrums created from the notes of a chord, we synthesise a chord waveform using a harmonic sinusoid model

$$y(t) = \sum_{j=1}^{J} \sum_{h=1}^{H} a(h) \sin(2\pi h f_0^j t) \quad (9)$$

where f_0^j is the fundamental frequency of the jth note in a chord and J is the number of notes in the chord. Given a synthetic signal as above, a chord template for a feature type is created by the same process in which the chroma itself is formed, calculating a (RA)-CQT and subsequent pitchgram and applying corresponding transforms. It is hoped, in this manner, to create a more realistic chord model. In particular, when compression is applied to the pitch feature, relatively large coefficients may be placed on e.g. sidelobe energy, which might be zero in an idealised spectral synthesis approach. Similar to the binary templates, a dictionary may be formed by cycling templates of a given class.

2.4. Chord estimation

As previously noted, chroma is a noisy feature in the presence of real audio signals, and the temporal continuity of the signal needs to be considered. Here, chord estimation is performed using a HMM-based classifier. The cosine distance, $b_{l,n} = \mathbf{t}_l^T \mathbf{c}_n$ is used as a measure of fit for the lth chord at the nth time frame. Quasi-probabilities are then calculated from this measure of fit by

$$Q_{l,n} = Q(\hat{l}_n = l) = e^{-\frac{1}{2\sigma^2}(1-b_{l,n})^2} \quad (10)$$

where $\hat{l}_n$ is the selected chord at the nth frame and σ is a user selected value. These quasi-probabilites are then input to the Viterbi algorithm. A simple transition matrix,$\mathbf{A} \in \mathbb{R}^{l \times l}$ in which all diagonal coefficients, relating self-transitions, are homogenous and all off-diagonals relating transition to another chord are also homogenous is employed here.

$$[A]_{i,j} = \begin{cases} \alpha/(1+N\alpha) & if \quad i \neq j \\ (1+\alpha)/(1+N\alpha) & if \quad i = j \end{cases} \quad (11)$$

where α is a HMM parameter that may be varied. This simple approach has been shown in several cases [7] [23] to result in similar results to a transition matrix that is derived from the probabilities of chord transitions which is attributed to the high probability of self-transitions. Employing this transition matrix can also been seen as using a change penalty term in the Viterbi cost function

2.5. Relationship to other work

Several of the steps were previously proposed for ACR, with some modifications applied here along with their use in different contexts. Reassignment-based chroma estimation was proposed in [15] which we modify here to directly compare to the CQT feature, and augment with the transforms of weighting, compression and filtering. We also consider the reassigned chroma with template models rather than the GMMs used in [15]. The synthetic template employed in [19] [6], is modified here by being generated from waveforms rather than spectral addition, in order to be used in the new context of enhanced chroma features. Finally, the chroma features were originally proposed in [24] [12], while we found that non-standard variants of CRP were preferable for template-based ACR in [13].

3. EXPERIMENTS

A dataset comprising the Beatles [25], Queen and Zweick [3] subsets of the Isophonics dataset; the RWC-POP [26] and US-POP [27] datasets which were annotated by Cho, and the Robbie Williams dataset [28], was employed. This dataset of 578 songs contains all data employed in [7] and [8]. Chord templates for major and minor chords were synthesised. All chords in the ground truth annotations were mapped to one of these classes, with minor and diminished chords mapped to minor and all other chords mapped to major, as is common practice for major-minor classification task. A silence detector was employed, similar to [7] with detected frames labelled as no-chords, a class that is also found in the ground truths.

Several chroma variants were calculated, including the basic chroma, C^A, log compressed feature, C^L, and CRPs with filter parameters of 15, 35 and 55, denoted in brackets. Weighted versions of these chroma features were also derived, denoted by C_W^A, C_W^L and WCRP for the basic, log, and CRP chroma features. Each feature was created using both the CQT and the RA-CQT. We consider the C_W^L with CQT as the baseline, as this was best in template-based ACR in [7]. The CQT was derived from code in [29], using windows of $186ms$ with $93ms$ overlap. The RA-CQT was implemented as described above, and was derived from initial spectrograms with a window size of $186ms$ and an overlap of $163ms$. Binary templates and the synthetic templates were compared. For the synthetic templates the number of harmonics was set to the different values $h \in \{4, 10, 20\}$ and the roll-off parameter was varied from $r \in [0.4 \;\; 0.8]$ in steps of 0.1. Analysis was performed the HMM-based classifier, which was run for a variety of values of α, the transition matrix parameter. Results given are for the optimal setting of α, h and r for each respective feature over all tracks in the dataset.

Each frame is labelled with a chord after classification, and the chord labels are compared to the ground truth. All frames are then labelled as true positive, $\mathcal{T}$ or false positive $\mathcal{F}$, allowing the recall metric to be calculated

$$\mathcal{R} = \frac{\#\mathcal{T}}{\#\mathcal{T} + \#\mathcal{F}} \times 100\%$$

	CQT		RA-CQT	
	Bin.	Syn.	Bin.	Syn.
$\mathcal{C}^A$	66.8	68.8	71.6	72.6
$\mathcal{C}^L$	65.9	69.5	73.4	75.0
CRP(15)	67.3	71.9	75.0	76.6
CRP(35)	66.8	72.3	74.6	76.7
CRP(55)	65.2	70.8	73.8	75.2
$\mathcal{C}^A_W$	70.8	71.2	74.1	74.5
$\mathcal{C}^L_W$	73.5	74.4	76.5	77.4
WCRP(15)	73.7	74.8	76.5	77.5
WCRP(35)	73.6	75.9	76.5	77.5
WCRP(55)	72.3	74.1	75.3	76.2

Table 1. Results for ACR experiments on the full dataset, comparing CQT with RA-CQT, binary and synthetic templates, and various chroma features using HMM-based classification.

3.1. Results

Results for the experiments are shown in Table 1. Here it is seen that the synthetic templates improve on the binary templates in all cases. In some cases, such as CRP(35) with the CQT, this results in improvements of over 5%. These differences are less when spectral weighting has been applied, which is to be expected as both the weighting and the synthetic chord model seek to reduce the effects of higher overtones. Likewise it is seen that the RA-CQT always results in improvements over the CQT for a given feature. This improvement relative to the CQT type is regardless of chord model i.e. binary templates with the RA-CQT improve on synthetic templates with the CQT. With binary templates, even the basic chroma, C^A with RA-CQT, performs better than all chroma with CQT, with respect to whether spectral weighting was employed. Spectral weighting itself is seen to improve ACR in all cases, as previously reported in [7] [13]. It is noticable however that this effect is muted when the RA-CQT is employed. In particular, applying weighting to CRP(15) and CRP(35) with synthetic templates and RA-CQT results in a difference of less than 1%.

Some improvements using the CRP relative to the log feature are observed when spectral weighting is not applied. However, unlike previously [13], we observe that the (W)CRP(15) performs better than (W)CRP(35) when the binary templates are employed, although the differences are small. We assume this is an effect of the more varied dataset used here, where we observe a more balanced selection of chords in terms of major-minor classes than in the dataset employed in [13]. However, with synthetic templates CRP(35) performs at least as well as CRP(15). With weighting applied, little difference is seen between C^L_W and WCRP(15) and WCRP(35), particularly with RA-CQT. [1]

[1] Code available at github.com/kooh7/RACHR

	Btls.	Iso	RWC	RW	[8]
$\mathcal{C}^L_W$	77.0	70.6	69.5	76.8	73.9
$\mathcal{C}^L_W$	79.8	78.6	77.4	79.4	78.3
WCRP(15)	79.9	78.6	77.3	79.6	78.3
WCRP(35)	80.2	78.7	76.9	79.6	78.3
DNN [8]	80.2	79.3	77.3	80.1	78.8

Table 2. Comparision of ACR performance Beatles (Btls.), Isophonics (Iso) RWC, and Robbie Williams (RW) and aggregate dataset from [8]. Top feature is the baseline approach from [7]. Centre three features use RA-CQT with synthetic templates. Bottom feature is DNN-based approach in [8].

Further results are given in Table 2, comparing the baseline, the best approaches from Table 1 and the deep chroma extractor [8] on the datasets from that paper [8]. A similar improvement over the baseline as in Table 1 is seen, while performance is close to that of the DNN chroma feature. The different approach used by the DNN to incorporate temporal context is noted. While a HMM is employed with the templates, the DNN is a convolutive network which uses several time frames to classify one frame. We note also the similar performance of our baseline and the baseline in [8], which uses logistic regression on multiple frames of the C^L_W feature.

To summarise, we observe several ways in which template-based ACR is improved, using reassignment, synthetic templates, spectral weighting, and using CRP(15) / CRP(35). However, a combination of these effects leads to sub-additivity in terms of improvements e.g. the log feature performs just as well as CRPs when weighting and RA-CQT are employed. The closeness in performance to the DNN chroma extractor may possibly suggest a saturation in terms of ACR performance for the minor-major ACR task in relation to feature extraction and simple temporal consideration.

4. CONCLUSIONS

We reconsidered template-based ACR, with reassigned spectrograms, synthetic templates and various chroma features. We found this combination effective, improving on the baseline feature by $\sim 4\%$, and performing similarly to a DNN-based chroma feature, whilst not requiring the expensive training of DNNs. This suggests much of the improvement in DNN-ACR is due to chord sequence modelling with recurrent neural networks (RNN) rather than the feature extraction. A simple HMM was employed here, guided by one parameter over all signals. Tuning the parameter song-wise results in $\sim 3\%$ improvement, while still not being optimal locally. Such local adaptivity is the strength of RNNs for ACR [30], and their use with the proposed features should be considered. We will also consider adaptive HMM models parameterisable from temporal features of a signal, which shall be compared to RNNs, as this may be more computationally attractive.

5. REFERENCES

[1] T. Fujishima, "Realtime chord recogntion of musical sound: A system using common lisp music," in *Proceedings of the International Computer Music Conference*, 1999, pp. 464–467.

[2] A. Sheh and D. Ellis, "Structured prediction models for chord transcription of music audio," in *Proceedings of ISMIR*, 2003, pp. 185–191.

[3] M. Mauch and S. Dixon, "Simultaneous estimation of chords and musical context from audio," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1280–1289, August 2010.

[4] J. Pauwels and J.-P. Martens, "Combining musicological knowledge about chords and keys in a simultaneous chord and local key estimation system," *Journal of New Music Research*, vol. 43, no. 3, pp. 318–330, 2014.

[5] K. Sumi, K. Itoyama, K. Yoshii, K. Komatani, T. Ogata, and H.G. Okuno, "Automatic chord recognition based on probabilistic integration of chord transition and bass pitch estimation," in *ISMIR*, 2008, pp. 39–44.

[6] H. Papadopoulos and G. Peeters, "Joint estimation of chords and downbeats from an audio signal," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 138–152, Jan 2011.

[7] T. Cho and J. P. Bello, "On the relative importance of individual components of chord recognition systems," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 477–492, Feb 2014.

[8] F. Korzeniowski and G. Widmer, "Feature learning for chord recognition: The deep chroma extractor," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2016.

[9] Y. Wu and W. Li, "Music chord recognition based on midi-trained deep feature and blstm-crf hybird decoding," in *ICASSP*, 2018.

[10] S. Sigtia, N. Boulanger-Lewandowski, and S. Dixon, "Audio chord recognition with a hybrid neural network," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2015.

[11] B. McFee and J. P. Bello, "Structured training for large-vocabulary chord recognition," in *ISMIR*, 2017.

[12] M. Müller and S. Ewert, "Towards timbre-invariant audio features for harmony-based music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 649–662, 2010.

[13] K. O'Hanlon, J. Pauwels, S. Ewert, and M. B. Sandler, "Improved template-based chord recogntion using the crp feature," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP)*, 2017.

[14] M. McVicar, R. Santos-Rodriguez, Y. Ni, and T. De Bie, "Automatic chord estimation from audio : A review of the state of the art," *IEEE / ACM Transactions on Audio, Speech and Language Processing*, vol. 22, no. 2, pp. 556–575, February 2014.

[15] M. Khadkevich and M. Omologo, "Reassigned spectrum-based feature extraction for gmm-based automatic chord recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 1, pp. 15, 2013.

[16] K. Lee and M. Slaney, "Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 2, pp. 291–301, February 2008.

[17] A. Weller, D. Ellis, and T. Jebara, "Structured prediction models for chord transcription of music audio," in *Proceedings of ICMLA*, 2009, pp. 590–595.

[18] Z. Rao, X. Guan, and J. Teng, "Chord recognition based on temporal correlation support vector machine," *Applied Sciences*, vol. 6, no. 5, 2016.

[19] L. Oudre, Y. Grenier, and C. Fevotte, "Chord recognition by fitting rescaled chroma vectors to chord templates," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2222–2233, Sept 2011.

[20] E. Gomez, "Tonal description of polyphonic audio for music content processing," *INFORMS Journal on Computing*, vol. 18, no. 3, pp. 294–304, 2006.

[21] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[22] K. R. Fitz and S. A. Fulop, "A unified theory of time-frequency reassignment," *CoRR*, vol. abs/0903.3080, December 2009.

[23] Meinard Müller, *Fundamentals of Music Processing*, Springer International Publishing, January 2015.

[24] M. Müller, S. Ewert, and S. Kreuzer, "Making chroma features more robust to timbre changes," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009, pp. 1869–1872.

[25] C. Harte and M. Sandler, "Automatic chord identification using a quantised chromagram," in *Proceedings of the Audio Engineering Society Convention*, 2005.

[26] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Rwc music database: Popular, classical, and jazz music databases," in *The 3rd International Conference on Music Information Retrieval (ISMIR)*, 2002, pp. 287–288.

[27] A. Berenzweig, B. Logan, D. Ellis, and B. Whitman, "A large-scale evaluation of acoustic and subjective music-similarity measures," *Computer Music Journal*, vol. 28, no. 2, pp. 63–76, June 2004.

[28] B. di Giorgi, M. Zanoni, A. Sarti, and S. Tubaro, "Automatic chord recognition based on the probabilistic modeling of diatonic modal harmony," in *8th international workshop on multidimensional systems (nDS13)*, 2013.

[29] T. Cho and J. Bello, "Large vocabulary chord recognition system using multi-band features and a multi-stream hmm," in *MIREX*, 2013.

[30] F. Korzeniowski, D. R. W. Sears, and G. Widmer, "A large-scale study of language models for chord prediction," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 91–95.