A Review of Modern Music Computational Chord Estimation Methods

Sean Bisset – 1905262

*Music* is a universal language that can be interpreted in many ways. Computational music analysis has been around for many years and is still a growing field. However, the analysis still needs to include many chords and concepts. For instance, most analysis methods do not consider chords other than major or minor, limiting the development of more complex Automatic Chord Estimation (ACE) models, therein further limiting deeper analysis of concepts such as emotion or quality of music. Some approaches also suffer from noise in the audio samples. This literature review aims to provide an overview of chromagram extraction methods and Audio Chord Estimation models, then critically analyse those methods and discuss methods of improvement.

Distinguishing key/chord/bass within a piece should be split into two stages: chromagram extraction from a piece and estimating the information based on the chromagrams.

Chromagrams or chroma features are a method of displaying audio frequency amplitudes across time. They are used most commonly when analysing audio in speech recognition and music analysis. Many method have been used to extract chroma features from sound files with varying success.

*Short-Time Fourier Transform (STFT)*Fujishima (1999) proposed that the Chromagram could be extracted from the audio signal by sorting each musical frequency into nearby frequency bins. These are then mapped onto one octave. This method is simple yet effective and has been used in many key/chord estimation models since the early 2000s (e.g., (M. Thomas et al. (2016), J. de Jesus Guerrero-Turrubiates et al. (2016)). This method, however, would suffer when noise is introduced into an audio sample as the frequency counts become distorted by the noise and the resulting chromagram would be hard to analyse. Noise is not the only problem; instruments that do not contribute to the harmony but instead infer the rhythm (percussive track) pose a problem as a chord cannot be derived from this. A different method,

Chart, application

Description automatically generated*Constant Q Transform (CQT).*Brown (1991) proposes that the Chromagram can be extracted by calculating the pitch energy class within the log-frequency domain. This method is derived from the knowledge that musical notes are equidistant in log-frequency. This method could then be used for instrument recognition using a pattern recognition algorithm. However, this method only works for music where the frequencies are spread equidistantly, as in Western music. Many non-western pieces would not work using the CQT method where they would in the STFT method, provided there is no noise or percussive track in the audio sample. Both the methods above also have the problem that chords are only sometimes beat synchronous. This would mean that the chromagram outputs are sensitive to noise and local transients (T. Cho and J. P. Bello 2011 p. 651). A problem that a few methods can resolve,

Figure - Example CQT Chromagram Output (Red – Higher Amplitude Frequency, Blue – Lower Amplitude Frequency

One of which is *Reducing Local Variations of Chords* (T. Cho and J. P. Bello 2011). The researchers proposed that repetition, something found continuously throughout music, can smooth the chromagrams and reduce noise. Smoothing the chromagrams is done by taking multiple similar sample chromagrams and averaging them out to smooth out noise and unsystematic deviations. This method of smoothing out noise is excellent. However, there would be a worry that it may remove interesting "one-off" parts that add to a piece. Many pieces add slight variations to the main thematic elements to create interest, and this method, if tuned incorrectly, could average these elements out. A simple example of this would be a piece where the V chord, G, is used in almost every bar, then to jazz up a section, the piece may make the G chord a G7 for a bar, which contains three of the same notes as the G chord and may therefore be averaged out. Instead, this could be fixed by ensuring averages are only taken locally instead of within the whole piece of music.

Another popular and simple method of filtering out noise in data is to simply pass the data through a median or low-pass filter. However, it has been found that using single pre-filtering methods alone, such as those listed above, does not yield a worthy impact on performance (T. Cho and J. P. Bello 2011 p. 651)

Returning to chromagram extraction, an improvement over the *CQT*method,

Harmonic Change Detection Function (HCDF) is a method proposed by C. Harte, M. Sandler, and M. Gasser. (2006) for detecting changes in the harmonic content of audio signals, which can be applied with the CQT method to produce better results. This is done by generating a 12-bin tuned chromagram with the log-frequency vectors output in the CQT method. Then, a harmonic Centroid transform is applied to the chromagram vectors; this is a 6-dimensional feature vector derived from tonnetz, which is explained deeper by Hu, Hanlin & Gerhard, and David. (2019). The harmonic centroid transform is a pitch representation showing pitch classes with similar harmonic relations. Imagining the 6-dimensional object as three harmonic circles, a circle of fifths; a circle of Major thirds; and a circle of Minor thirds, helps to understand how distance is measured within the 6D space. Smoothing is then applied in the way of a gaussian filter to reduce noise before the distance calculation is made to determine Harmonic Change. This 4-step process is known as the HCDF. Preliminary experiments to detect chord changes gave a precision score of 64.9%.

Diagram, radar chart

Description automatically generated

Figure - Visualising the 6-D Tonal Space as three circles. The Tonal Centroid for chord A Major is shown at point A (C. Harte, M. Sandler, and M.Gasser 2006)

Timbre-Invariant Audio Extraction (TIAE) proposed by M. Muller and S. Ewert (2010), is a method of returning timbre-separated chromagrams for easier feature extraction. This method exploits the difference between octaves that different timbres have to separate the output into easier-to-analyse chroma features. The results show that using this method can significantly increase feature extraction models' accuracy. The work done here is excellent, as many estimation models suffer from noise of other timbre and overlapping frequencies. The method can still be improved by reducing the number of pitch-frequency cepstral coefficients (PFCC) since the experiments in the paper revealed that this could be done without impacting the output.

The problem percussive tracks create when using these methods can be negated by separating the audio into separate tracks and chromagrams. Some strategies for achieving this are listed below:

*Harmonic/Percussive Sound Separation (HPSS),* proposed by Ono, et al. (2008), that the harmonic and percussive tracks could be separated easily since harmony usually has a well-defined frequency content. In contrast, percussive parts are typically spaced in time to create the piece's rhythm. This method is excellent at separating music with a well-defined rhythm but would struggle if the rhythm were inconsistent or polyrhythmic. Ono, et al. (2008) then continued to show that the output chromagrams from this method only yielded better chord estimation performance.

**Chord Data Extraction from the Chromagrams**

Getting the chord data from the chromagram is a complex task. The most common approach for extracting the chord sequence is to use Hidden Markov models (R. Rabiner, 1989). L. E. Baum and T. Petrie (1966) proposed the HMM as a probabilistic model where the sequence modeled is considered a Markov process of a sequence of related hidden variables. In the case of Chord Estimation, the desired Chords are the hidden variables, whereas the chromagram/chroma features are considered the observed variables. This method requires an expert to tune the HMM variables. It is, therefore, less accessible and understandable. A. Sheh and D. P. W. Ellis (2003) proposed that the HMM could be tuned using an expectation maximisation approach. The EM algorithm allows hidden variables to be included in classifier parameter estimation. As in ML, tweaking the HMM is iterative until the model is optimal or "trained." This trained model then goes on to estimate chords based on the chromagrams. The method is, again, excellent. However, it should be noted that a labelled dataset should be supplied for the training of the HMM. Since a dataset is provided for training, it may be worrisome that the model may attune to a particular genre of music since many genres have different time signatures and features. A. Sheh and D. P. W. Ellis (2003) found that when trained with eighteen songs from three early Beatles albums and tested with two songs, the model produced recognition accuracies below 30% and alignment accuracies majority less than 50%. These accuracies display the potential that chord estimation methods have for improvement.

Chart, scatter chart

Description automatically generated Another more recent development by A. Weller, D. Ellis, and T. Jebara (2009) uses a structured support vector machine (SVMstruct) algorithm to attempt prediction on a similar dataset of Beatles songs. However, the SVM method becomes more accurate than the HMM solution only when the dataset size increases. The SVM method above is promising; however, a significant push can still be made to improve the accuracy and prediction time. Figure three shows the results of multiple tests on different SVM models and a HMM. It can be seen that the SVM approach becomes more accurate as the quantity of data used for training is increased. becomes more accurate as the quantity of data used for training is increased.

Figure - Results from multiple SVM model trained on different quantities of data. (A. Weller, D. P. W. Ellis, T. Jebara 2009)

*Dynamic Bayesian networks* are a more complex HMM model topology. The topology considers the key and metric position of the Markov to predict the chord probabilistically. This method, however, is not fully autonomous and requires expert tuning but has been proven superior to the less complex HMM model approaches discussed previously. As aforementioned, having an expert required to fit the network before usage is not attractive or accessible. A more autonomous system would allow for faster and more large-scale chord estimation. This method, therefore, is unfortunately not suitable for that purpose.

In 2012, Y. Ni, et al. (2012) proposed the *Harmonic Progression Analyser (HPA)*, an End-to-End machine-learning solution for chord, key, and bass estimation. The researchers presented a novel Chromagram extraction method as well as a novel machine-learning method of chord prediction. The HPA method was incredibly successful in Chord Estimation, also showing outstanding accuracies in complex chord estimation. The biggest drawback of the system, however, was the processing time. This could be improved by writing the chord estimation algorithm in a faster programming language such as C or C++. The paper also noted that some chord estimations were found to be incorrect due to the bass Markov chain "overpowering" the chord Markov chain.

A paper by N. Yasui, M. Miura and T. Shiumamura (2019) showcases a novel method that considers the difference in Electric Guitars. This is done by converting the sounds of electric guitars into chroma vectors representing the power of all pitch classes for a twelve-tone scale. Then the system attempts to recognise notes performed and assign chord labels based on the recognised chroma vectors. The research results showed that the differences between the guitars are reduced when the method is implemented, making chord extraction more straightforward and accurate. The method is very intriguing because it shows that instrument differences can cause a different prediction to be made. Furthermore, it proves that accuracy continues to suffer from the variety of instruments found within modern music.

In 2020 a semi-supervised neural-network-based approach was devised by Y. Wu, et al. (2020). The approach combined the research of past HMM solutions with the development of Artificial Neural Networks to create multiple systems with high accuracies. It could define slightly more complex chord notations. This enormous development continues to show that past research helps to forward more impressive and complex systems. The accuracy reached 82.80%, showing that a trained Neural Network method is excellent for Automatic Chord Estimation models. Furthermore, the paper concluded that further research should be done to develop a more comprehensive system capable of predicting other aspects of music. Therefore, showing that chord estimation model research can be applied in other audio analysis fields, and any research done will improve the overall understanding of music, audio, and speech recognition.Table

Description automatically generated

Figure - An example of chord label sequences estimated by the supervised and semi-supervised methods without the Viterbi post-filtering. For readability, only the first 24 dimensions (bass and middle channels) of the chroma vectors are displayed. (Y. Wu, et al. 2020)

Chordify (2021) is an online Chord transcription service that allows users to input a song that the model should analyse. In moments, the system can return the chords of a song. However, the chords returned are not always accurate. Furthermore, Chordify only selects chords from a subset of 24 basic chords, missing out on conveying the beautiful nuances that more complex chords offer. Chordify also requires a premium subscription to be paid monthly in order to transcribe multiple songs a day. A neural network (NN) implementation is used on Chordify. However, not much information is given on the model's exact specifications. The Chordify NN is trained on a database of songs often. But, as aforementioned, the system misses out on a vast range of chords and is, therefore, not at the potential ACE models could be. Surprisingly, Chordify cannot predict complex chords, given that Y. Wu, et al. (2020) developed a system capable of more complex chords than simply the 24 Major and Minor that Chordify offers. Chordify CEO Bas de Haas stated, "In the future we hope to be able to offer you chords that fit your level of playing. Another idea is to prioritize chords that you use more often over more standard chords. The options are endless, really." (Chordify 2021), another example of the vast amount of work to be done for audio analysis and Chord Estimation methods.

To conclude, the purpose of this review was to analyse the current and previous methods of audio analysis to reveal the vast amount of work that is to be done in order to achieve a deeper analysis of music and audio. The research has shown the improvement still to be made in Chromagram Extraction methods and Audio Chord Estimation models. The research has also clearly shown that methods are still imperfect and do not return the more complex chords that make music an exciting soundscape full of emotion and intrigue.

**Requirements Analysis**

Functional Requirements:

* All Users:
  + - Users must be able to input a piece of music.
    - Users must be able to view the chords of input songs.
    - Users must be able to install the application on a windows system.
    - Users must be able to uninstall the application.
    - Users must be able to delete any songs they have uploaded.
    - Users must be able to consent to data collection for further development.
    - Users should be able to edit the chords of a song.
    - Users should be able to play a song within the app.
    - Users should be able to pause a song within the app.
    - Users should be able to view the chroma features generated for a song.
    - Users should be able to select regions of a chroma feature for re-prediction.
    - Users should be able to stop analysis in progress.
    - Users should be able to use the application offline
    - Users should be able to see file name of imported files
    - Users should be able to see file path of imported files
    - Users should be able to play a song with only the recognised chords
    - Users should be able to view the audio as an amplitude transform
* Application:
  + - Must be free to download and install
    - Must contain a navigation menu
    - Must run smoothly without crashes
    - Must be able to install/be used on a windows device
    - Should store user preferences

Non-Functional Requirements:

* Interface/Ui Design:
  + - The application should be neatly organised and simple.
    - The application should contain a light and dark mode
* Security:
  + - The application should allow users to consent to data collection
    - Data that could be taken should be stored and transferred securely
* Maintainability:
  + - The application should be written with an Object-Oriented Programming method
    - Code files should be well documented and clear
* Data Collection:
  + - The application might store data in an API for training
    - The application might allow a user to submit training data to the API for model training
    - The application should disclose data collection methods to the user
* System Requirements:
  + - The application should run on windows devices above the minimum specification
    - The application could be able to run on low-end systems
    - The application should run without frequent crashes
* Support:
  + - Users must be able to submit issues to developer via application interface
    - The application should offer the user updates when available
* Accessibility:
  + - The application should be easy to download from the internet
    - The application should be easy to install and run
    - Updates to the application should be done when the user accepts the update

A Review of Modern Music Computational Analysis Methods

References

A. Harte, C., Brian Sandler, M. & Gasser, M., 2006. Detecting harmonic change in musical audio. *AMCMM '06: Proceedings of the 1st ACM workshop on Audio and music computing multimedia,* 1(1), pp. 21-26.

Brown, J. C., 1991. Calculation of a constant Q spectral transform. *The Journal of the Acoustical Society of America,* 89(1), pp. 425-434.

Cho, T. & P. Bello, J., 2011. *A Feature Smoothing Method for Chord Recognition Using Recurrence Plots,* New York City: New York University.

de Jesus Guerrero-Turrubiates, J. et al., 2016. *Guitar audio signal classification by collapsed Pitch Class Profile.* Ixtapa, IEEE.

Fujishima, T., 1999. *Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music.* Stanford: ICMC.

Hu, H. & Gerhard, D., 2019. *Modelling 4-dimensional Tonal Pitch Spaces with Hopf Fibration.* Merseilles, Department of Computer Science, University of Regina.

Muller, M. & Ewert, S., 2010. Towards Timbre-Invariant Audio Features for Harmony-Based Music. *IEEE Transactions on Audio, Speech, and Language Processing,* 18(3), pp. 649-662.

Ono, N., Kameoka, H., Le Roux, J. & Sagayama, S., 2008. *Separation of a monaural audio signal into harmonic/percussive components,* Tokyo: The University of Tokyo.

R. Rabiner, L., 1989. A Tutorial on Hideen Markov Models and Selected Applications in Speech Recoginition. *Proceedings of IEEE,* 77(2), pp. 257-286.

Sheh, A. & P. W. Ellis, D., 2003. *Chord Segmentation and Recoginition using EM-Trained Hideen Markov Models,* s.l.: International Symposium on Music Information Retrieval.

Thomas, M. et al., 2016. *Detection of similarity in music files using signal level analysis,* Karnataka: Deapartment of CSE, National Institute of Technology Karnataka,Surathkal, Mangalore, India.

Weller, A., P. W. Ellis, D. & Jebara, T., 2009. *Structured Prediction Models for Chord Transcription of Music Audio.* Miami Beach, Columbia University.

Witte, J., 2021. *AI-technology behind the chords of Chordify – our algorithm explained.* [Online]   
Available at: https://chordify.net/pages/technology-algorithm-explained/  
[Accessed 05 11 2022].

Wu, Y., Carsault, T., Nakamura, E. & Kazuyoshi, Y., 2020. Semi-Supervised Neural Chord Estimation Based on a Variational Autoencoder With Latent Chord Labels and Features. *IEEE/ACM Transactions on Audio, Speech, and Language Processing ,* Volume 28, pp. 2956-2966.

Yasui, N., Miura, M. & Shimamura, T., 2019. *Chord Label Estimation from Acoustic Signal Considering Difference in Electric Guitars.* Taipei, IEEE.

Baum, L. E. & Petrie, T. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains." Ann. Math. Statist. 37 (6) 1554 - 1563, December, 1966. https://doi.org/10.1214/aoms/1177699147

A Review of Modern Music Computational Analysis Methods

Bibliography

Lee, K., 2008, “A system for acoustic chord transcription and key extraction from audio using hidden Markov models trained on synthesized audio,” Ph.D. dissertation, Stanford Univ., Stanford, CA

Noland, K. & Sandler, M., 2006 “Key estimation using a hidden Markov model,” in Proc. ISMIR

K. Lee and M. Slaney, “Acoustic chord transcription and key extraction from audio using key dependent HMMs trained on synthesized audio,” IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 2, pp. 291–301 Feb. 2008.