Revised Selected Papers

Accademia Musicale Studio Musica
Michele Della Ventura, *editor*

2024

Proceedings of the International Conference on New Music Concepts

Vol. 11

ACCADEMIA MUSICALE
STUDIO
MUSICA

Accademia Musicale Studio Musica

International Conference on New Music Concepts

Proceedings Book
Vol. 11

Accademia Musicale Studio Musica
Michele Della Ventura
Editor

Published in Italy
First edition: April 2024

www.studiomusicatreviso.it
Accademia Musicale Studio Musica – Treviso (Italy)
ISBN: 978-88-944350-5-4

Automatic Creation of Chordal and Melodic Chromaticism

A. Christou, R. Dannenberg

Yale University | Carnegie Mellon University
antonis.christou@yale.edu | rbd@andrew.cmu.edu

Abstract. Western popular and art music use the diatonic scale as their foundation in both melody and harmony, and restricting to diatonic scales can simplify both analysis and composition. However, even mostly diatonic genres contain some limited use of chromaticism or accidentals. Therefore, we propose that, after creating diatonic music, we can automatically alter selected notes chromatically to enhance the melody and harmony. This paper presents several deep learning models for generating this chromaticism given a purely diatonic input melody and accompanying chord sequence. The altered melody and harmony from the model are generally compelling and conform to musical syntax, making our work a potentially useful tool for both human and computer agents to augment existing compositions or aid in the compositional process.

Keywords. Chromaticism, Machine Learning, Melody, Music Generation and Composition, Trans former

1 Introduction

The concepts of diatonicism and chromaticism are fundamental building blocks in Western music. Diatonicism is generally seen as the most basic harmonic system and refers to the use only of notes from a particular diatonic scale. For example, a song in the key of C major that exclusively uses the notes of that key would be seen as purely diatonic. In contrast, the chromatic scale refers to a division of the octave into twelve equal half steps and does not fit neatly into any one key.

Fig. 1. A C major scale.

While the use of notes and derived chords from diatonic scales is most common, composers from all eras occasionally make alterations to the notes of this diatonic scale and instead use notes and chords from the chromatic scale. This process is called chromaticism and idiomatic use of it is fundamental to the unique sound of many genres of music [1, 2].

Therefore, due to the great importance of using chromaticism effectively in writing compelling music in many genres, developing models that generate or harness it in their compositions is an important but largely unexplored area up until this point. The most common approach has been to create a model that learns from data when to use chromatic and diatonic notes and so freely selects notes not restricted to the diatonic pitches. However, this places the burden on the model of learning the difference between the two sets of pitches and often results in outputs that drift from one key to another or includes a number of "wrong" notes due to the probabilistic nature of these methods.

Fig. 2. A chromatic scale.

As a result, some successful automatic composition models are restricted exclusively to the diatonic scale, eliminating the possibility of incorrect chromatic notes but also resulting in simpler sounding outputs. Our goal is to create a tool powered by deep learning that could potentially be used in conjunction with existing diatonic models or human compositions. To that end, we propose a "post-processing" or editing step that inserts chromatic melody notes and chords. We believe that this model breaks new ground in this area of deep learning applied to musical chromaticism and will enable future explorations as well as lead to generally more interesting sounding music generations now and in the future.

2 Related Work

Even before the invention and widespread use of neural methods, composers ranging from Mozart with his games of dice to Xenakis and Cage's aleotoric music have used algorithmic or computational methods to produce musical output [4]. Today, a great deal of work has been produced on the subject of creating models to generate music using a variety of differ ent methods. Markov chains and grammars were used both historically and, in the present, producing notable works such as Hiller and Isaacson's

Illiac Suite, which is one of the earliest scores produced by a computer [6, 7]. More recently, recurrent neural networks have been applied, recognizing the inherently sequential nature of music [5]. Other methods include convolutional neural networks, notably applied by Huang et al. [9] to the problem of completing musical scores in the domain of contrapuntal polyphony. Most recently, the Transformer with its self-attention and ability to encode long term structure [8] in addition to other deep learning methods have been applied to music generation with promising results [9]. Work has also been done in the area of harmonizing melodies with appropriate chord changes [10]. However, the problem of chromatizing existing diatonic melodies and harmonies rather than simply generate them has not been explored previously, despite its great importance in musical style. A variety of models have used the simplifying assumption of generating purely diatonic outputs, ranging from rule-based systems [11, 12], to a deep-learning system [13], opening the door for the creation of a model that could be used to augment other exclusively diatonic compositions. Therefore, we introduce this model as a means of exploring this area and enabling human and computer composers to improve the quality of their material.

3 Data

A primary challenge in creating a model to learn patterns of melodic and chordal chromaticism is finding a data source with sufficiently rich harmonizations. We use the Public Domain Song Anthology book (PDSA), a collection of 347 songs taken from popular American folk songs that usefully for this application are provided with two sets of harmonizations, one slightly more diatonic and one using more advanced and modern jazz harmonies.

As a pre-processing step, we filter the PDSA to use only the songs in simple duple meter (ex. 2/4 and 4/4). Additionally, effort was taken to create a purely diatonic input to train with by eliminating all accidentals in the melody and all chords outside of the basic diatonic ones, leaving us with a chordal vocabulary of the 14 chords found diatonically in the keys of C major and minor for input and the 194 chords found in the PDSA with their chromaticism intact for output. The melody notes are encoded with midi as tuples in the form of [midi note, duration] with 0 being used to mean rest leading up to the highest midi note found in the PDSA, 87 or Eb6. Each song overall is encoded as a stream of these tuples for both melody and the attached chords. We also transpose all songs to the key of C major or minor for easier learning.

Fig. 3. An example song from the PDSA.

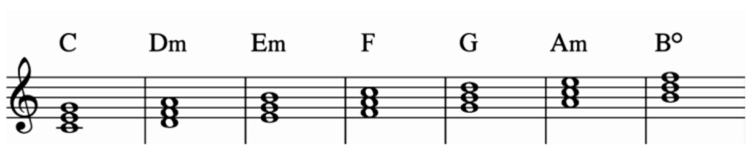

Fig. 4. The chords diatonic to C major.

4 Model

Our model uses both Transformer and LSTM, with the Transformer functioning as the encoder and LSTM as the decoder. This Transformer-LSTM hybrid model has been used successfully in other recent music generation projects [13]. We also tried several other models such as pure LSTM, pure bidirectional LSTM, and pure Transformer. We use a nested model to handle the melodic and chordal chromatization in two passes. For the chords, we encode each chord name as an integer in the vocabulary and pass this input to an encoding layer before going to the Transformer encoder. This representation is then decoded by an LSTM before finally going through a fully connected layer to generate the next predicted chord based upon the output chromatic chord vocabulary. After all chords have been processed, this process is then repeated with the melody notes. Each note encoded as midi is passed through a melodic model with the same architecture to make the binary classification decision of whether or not this melody note should be chromaticized or should remain diatonic. We use the convention of chromatizing notes based upon their most likely alteration in the C major scale (ex. Fs become F#s rather than Gbs, As become Bbs rather than A#, etc) as a simplification. After both models pass on the input, we recombine the newly chromaticized chords and melody with the durations found in the original song and generate a MusicXML file for viewing as a score according to the conventions of Western notation.

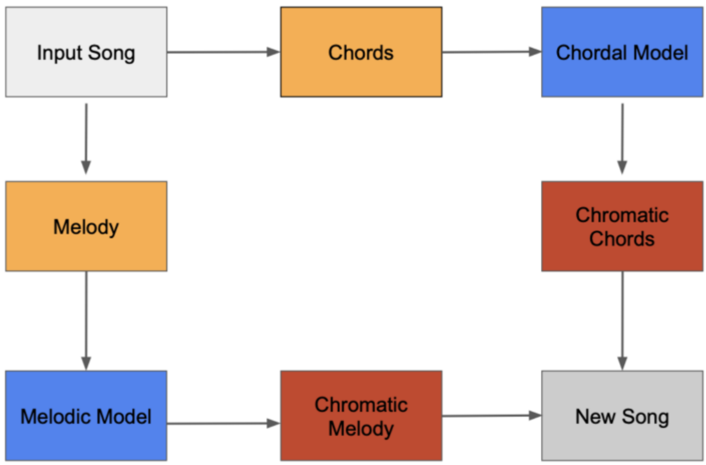

Fig. 5. Generation pipeline for new chromatic song.

This score includes both block voicings of the new chords as well as chord symbols in lead sheet notation, allowing for human musicians to play, alter, and voice the generations in addition to hear them as they are.

5 Experiment and Evaluation

To determine the best model for this problem, we first compared the four architectures above to see which produced the most accurate and idiomatic chromaticism. Melodic accuracy was defined as the percent of correctly made chromaticization decisions, while chordal accuracy was defined as the percent of correctly chosen output chords across the dataset given any particular input chord.
The models were trained using the songs from PDSA with 80% being reserved for training and the remainder for validation. Since the data for melodic chromaticism is already heavily unbalanced, with approximately 90% of the notes being diatonic, we opt to filter out any songs that do not contain any instances of melodic chromaticism for training the melodic model. We also set an optional flag for filtering out minor key songs to make for easier learning. After all filtering, 176 songs were used to train the

chordal model and 121 for the melodic. We use Cross Entropy Loss as our loss function and Adam as the optimizer. For the melodic model, as a means of combating the aforementioned unbalanced class distribution, we weight the loss function according to the prevalence of diatonic versus chromatic notes in the original dataset. In all networks, we use a learning rate of 0.001 and a dropout probability of 0.1. The transformer based models used two layers each for their respective encoder and decoders, with two heads in the multi-head attention in the self-attention layer, an embedding dimension of 80, and a feed forward dimension of 400. The LSTM component of the models that include one utilizes a hidden dimension of size 100.

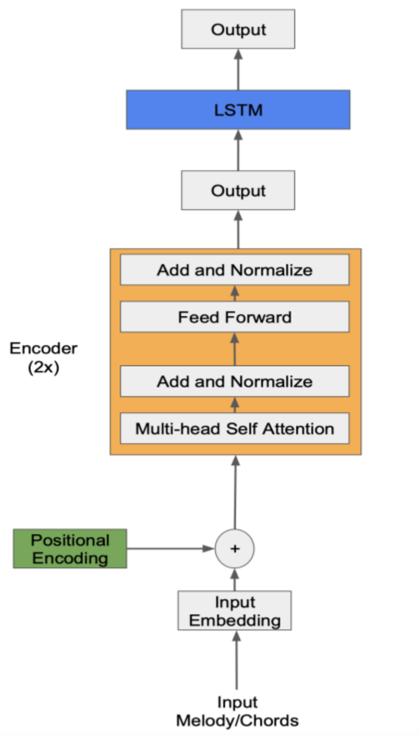

Fig. 6. Transformer-LSTM model used for melody and chordal chromaticization.

The pure LSTM model was augmented with a few hands selected features based upon domain knowledge, such as the duration of the melody note, the duration of the chord, and which chord generated earlier is playing over any given melody note, as these features are important to human musicians when deciding how, if at all, to chromaticize a note or chord. These values are concatenated with those from the embedding layers to form the full input to the model

Model	Melodic Accuracy	Melodic Loss	Chordal Accuracy	Chordal Loss
LSTM	68.5%	0.53	69.5%	1.53
Bidirectional LSTM	79.6%	0.69	72.2%	**1.36**
Transformer	72.1%	0.74	45.8%	3.63
Transformer-LSTM	**80.3%**	**0.36**	**73.16%**	1.47

Fig. 7. Loss and accuracy on validation set for each model type

Overall, the Transformer-LSTM model had the highest accuracy on the validation set. The Trans former model performed poorly as it tended to overfit heavily on the validation set and also struggled to train, likely because of the small dataset. For the LSTM based models, making the model bidirectional improved the accuracy, likely due to the fact that knowing what comes both before and after a given note or chord can help lead to idiomatic and grammatical chordal and melodic patterns.

6 Conclusion

Overall, the outputs from the model can at times be quite coherent and pleasant to listen to.

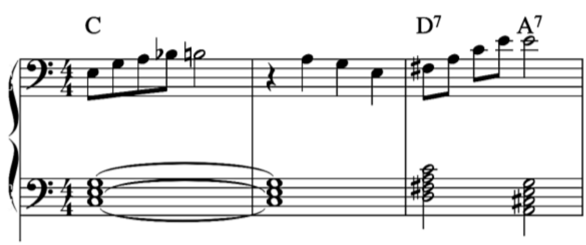

Fig. 8. An output phrase with proper use of chromatic passing tones and chromatic chords according to standard jazz harmony conventions and the PDSA.

The short example phrase above from the validation set correctly learned to use common chromatic melody devices like passing tones, and the generated chord progression uses a number of grammatical and typical stylistic hallmarks of chromatic harmony, such as secondary dominants and seventh chords.

Fig. 9. An example of dissonant, unidiomatic output in the melody.

However, at times some of the generations from the model can be dissonant and not grammatical according to the conventions of the styles represented with the PDSA. In the following example [Fig. 9], the model has outputted a chord and melody sequence that is theoretically plausible in isolation, but the G# in the melody voice directly clashes with the root of the chord, making it an unlikely choice for a human composer to write. Therefore, in future explorations, we would like to experiment more with allowing the model to more consistently select chords and melody notes that do not clash with each other, as well as investigate further into variable length and Seq2Seq based model, especially for melodic chromaticism.
The model in its current form can be used by humans and computer music models as a tool to give new ideas or make existing compositions more exciting. We believe that this model breaks new ground in a mostly unexplored area of music generation and opens the door for further investigation in this field.

References

[1] Meyer, Leonard B. (1956). Emotion and Meaning in Music, p. 217. Chicago: University of Chicago Press. ISBN 9780226521374. Quoted in Brown (1986), p. 1.
[2] Forte, Allen, Tonal Harmony, third edition (S.l.: Holt, Rinehart, and Wilson, 1979): p. 4. ISBN 0-03-020756-8 3.
[3] Cooper, Paul (1975). Perspectives in Music Theory, p. 208. New York: Dodd, Mead, and Company. 4.
[4] Taruskin, Richard. “Music in the Late 20th Century, Chapter 2 ‘Indeterminacy.’ ” The Oxford History of Western Music, Oxford University Press, Oxford Etc, 2010.
[5] Chen, Chun-Chi J., Miikkulainen Risto, 2001, Creating Melodies with Evolving Recurrent Neural Networks, IJCNN-01

[6] Thornton, C. (2009). Hierarchical markov modeling for generative music. In International computer music con ference.

[7] Lejaren A Hiller Jr and Leonard M Isaacson. Musi cal composition with a high speed digital computer. In Audio Engineering Society Convention 9. Audio Engi neering Society, 1957.

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. In Advances in neural infor mation processing systems (pp. 5998–6008).

[9] Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., . . . Eck, D. (2018). Music transformer. arXiv preprint arXiv:1809.04281 .

[10] S. Rhyu, H. Choi, S. Kim and K. Lee, "Translat ing Melody to Chord: Structured and Flexible Harmo nization of Melody With Transformer" in IEEE Ac cess, vol. 10, pp. 28261-28273, 2022, doi: 10.1109/AC CESS.2022.3155467.

[11] A. Elowsson and A. Friberg, "Algorithmic composi tion of popular music," in the 12th International Con ference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, 2012, pp. 276–285.

[12] Dai, Ma, Wang, and Dannenberg, "Personalized Pop ular Music Generation Using Imitation and Structure," Journal of New Music Research, (2023).

[13] Dai, Jin, Gomes, and Dannenberg, "Controllable Deep Melody Generation via Hierarchical Music Structure Representation," in Proceedings of the 22nd International Society for Music Information Retrieval Conference, Online, Nov 2021, pp. 143-150.

This book presents a collection of selected papers that present the current variety of all aspect of music research, development and education, at a high level. The respective chapters address a diverse range of theoretical, empirical and practical aspects underpinning the music science and teaching and learning, as well as their pedagogical implications. The book meets the growing demand of practitioners, researchers, scientists, educators and students for a comprehensive introduction to key topics in these fields. The volume focuses on easy-to-understand examples and a guide to additional literature.

Michele Della Ventura, editor

New Music Concepts and Inspired Education

Revised Selected Papers

ISBN: 978-88-944350-9-2

www.studiomusicatreviso.it