Neo-Riemannian Theory for Generative Film and Videogame Music

Sara Cardinale[1] and **Simon Colton**[1,2]
[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
[2]SensiLab, Faculty of Information Technology, Monash University, Australia

s.cardinale@qmul.ac.uk s.colton@qmul.ac.uk

Abstract

Music is an essential element of films and videogames, which strongly contributes towards an immersive experience, by establishing a setting and mood, enhancing the storyline, and developing characters. Automatically generating music for films and games has been explored in existing works, but there is still room for improvement in terms of musicality and adaptivity. Neo-Riemannian Theory (NRT) comprises an established set of techniques for analysing music which is triadic but not necessarily tonal, and has had much application to the study of emotional and situational arcs in film music. NRT has barely been used in a generative setting, and we introduce a rationalised version for this purpose, which we believe could have particular application to film and videogame music, where mood or emotion-based music is required. We suggest ways in which such a procedural NRT approach could be applied, and describe future directions for our own projects in this area.

Introduction

Music is central to the setting of ambiance, mood and emotional arcs in films and videogames, and often contributes to the portrayal of plot elements and character development. Generative music can be employed for offline film/game development and also in response to player actions and environment changes during live gameplay (Plut and Pasquier 2020). The adaptive nature of such generative music can often be beneficial, e.g., when experiencing music generated by the *Adaptive Music System* (AMS) (Hutchings and McCormack 2020), gamers reported "... an overall higher immersion and correlation of music with game-world concepts with the AMS than with the original game soundtracks ..." However, existing generative music systems for games can be said to lack somewhat in *musicality*, i.e., the ability to present music in a pleasing way, take sounds and arrange them in patterns and phrases using music theory concepts such as rhythm, harmony, dynamics, tone, articulation, form, musical continuity, and tempo to express thoughts and emotions. For instance, in (Hutchings and McCormack 2020), the authors state that: "listening to the music tracks, it becomes apparent that the overall music quality of the AMS is not that of a skilled composer ... musical quality could be enhanced through improvements to the composition techniques within the framework presented." Moreover, Liapis et al. state that there is not a substantial difference in the generation of game audio and music generation outside of games (Liapis, Yannakakis, and Togelius 2014).

In general, the systems used in industry are usually game specific, while systems made for academic research are too general (Plut and Pasquier 2020). For instance, the generative compositional system for *The Audience of Singular* videogame by Plut (2017) produces music which is not entirely adaptive and could be considered as an independent musical piece (Plut and Pasquier 2020). For melody generation, the system uses Markov Models to generate various possible music sentences or notes that are randomly selected from a scale. Such an approach does not follow organizational elements such as pitch proximity and late-phrase declination (Huron 2006) and can result in lack of musicality. Similar limitations can be found in generative systems for film music, such as in the *DeepScore* project (Savery and Weinberg 2020), where music for a film clip can be generated to fit temporal keyword cues, or to fit a visual analysis of the clip. Here, an evaluation of the system resulted in feedback that if the system used keywords instead of visual analysis of the media, the generated music was pleasant to listen to, but was often not in-sync with the on-screen actions, and the score was too simplistic and lacking emotion. A similar response was given for the system when using visual analysis, with additional feedback opining that transitions between scenes were too unharmonious.

We are developing a general generative system which can adapt to the requirements of a scene from a specific film or game to produce music with high musicality and appropriate support for the scene. To this end, we propose a procedural version of an established music analysis technique called Neo-Riemannian Theory (NRT). This has been used to successfully analyse emotional and situational arcs in film music, especially where *triadic* but not necessarily tonal music has been composed. With the term major/minor triadic chord (or trichord), we mean a chord of three notes with a tonic note, another note a major or minor third above the tonic, and a final note a fifth above the tonic. These can be presented in various permutations, or *inversions*. Our procedural version of NRT comprises a set of rewrite rules for minor and major trichords, and inherits a mapping from chord progressions to changes in emotional and situational elements in media such as films or videogames. Before describing NRT and procedural NRT, we first describe the roles that music plays in films and games. We end by describing related work and future directions for our work.

The Roles of Music in Films and Videogames

Films put demands on the music for their soundtracks that are quite different from those imposed on other genres (Lehman 2018), with videogames perhaps being the closest in this respect. In general, soundtracks must take on various roles to enhance and accompany the media's narrative. Copland delineates five purposes of film music: creating a convincing atmosphere of time and place; underlining psychological refinements (the unspoken thoughts of a character); serving as a kind of neutral background filler; building a sense of continuity; and underpinning the theatrical build-up of a scene, eventually rounding it off with a sense of finality. In film music, while the basic building blocks are the same as any other Western genre, how these are arranged is guided by the script, where the end goal of the music is making meaning (Lehman 2018), performing a narrative function opposed to solely being musically pleasing.

In order to further understand how soundtracks are composed to fit the video, Lehman (2018) sets out three Hollywood practices: (a) the soundtrack's active role in making meaning, (b) the tendency for film music to rely on immediate gestures for expressive impact where tonality is not necessary, via the practice of using small musical ideas chained together, repeated or changed to prevent the music from being overwhelming, by constantly giving the audience new musical information, and (c) the music's ability to make meaning via associations e.g., linking a feeling to a melody and being able to recall it through repetition.

Soundtracks for films and videogames share many functionalities. In both cases, timed musical cues and sound effects typically suggest a responsive, narrative-specific environment aimed at either immersing the viewer/player in the spectacle of storytelling or engaging them in the bodily emulation of problem solving in a narrative-based context. These principles give music its ability to create a compelling and entertaining emulation, as described specifically for game music in (Whalen 2004). Moreover, videogame music draws from various cinematic practices to help structure the game's narrative elements based on familiar dramatic conventions (Rod 2007).

Neo-Riemannian Theory for Musical Analysis

NRT originated in the work of David Lewin (1982), and was built on Hugo Riemann's work on interrelations of triads and systems. An example of this is harmonic dualism (negative harmony), which describes the inversional relationship between major and minor chords, with minor triads being considered "upside down" or mirrored versions of major triads (Rehding et al. 2003). NRT was a response to analytical problems created by chromatic music that is triadic but not necessarily tonal (*triadic chromaticism*) (Lewin 1982). Analytical models for diatonic music are not suitable to analyse chromatic chord progressions, as chromaticism makes use of notes foreign to 7-note modes or diatonic scales (Cohn 1998). Therefore, Neo-Riemannian Theory has been used as an analysis tool for chromatic chord progressions, including in film music, showing that tonality is not the only way to relate chords.

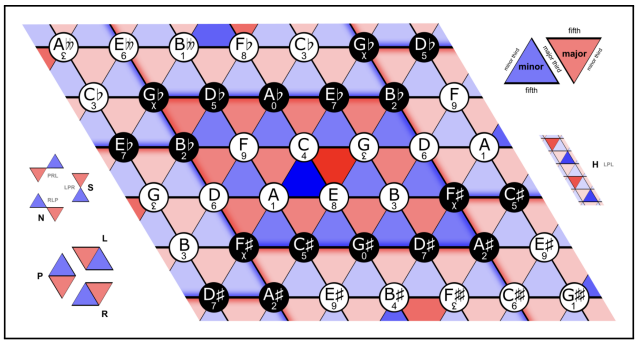

Figure 1: The Tonnetz, from en.wikipedia.org/wiki/Tonnetz.

NRT employs three Neo-Riemannian Operators (NROs) to describe the transition from one trichord to another. These are: *Parallel (P)* for pairs of triads that share an interval of a fifth, *Relative (R)* for triads that share a major third, and *Leading-tone Exchange (L)* for triads that share a minor third. When analysing music, the chord transitions which can be identified with these operators are dictated by the *Tonnezt* depicted in figure 1, which has been used since the 18th Century to describe chord progressions. To describe the P, L and R operations, we take a triangle on the main board of either red or blue colour and map it to the adjacent triangle, as per the key in the bottom left of figure 1. Red triangles represent major trichords and blue triangles represent minor trichords, and we see that P, L and R always transform major to minor chords and vice-versa.

Part of the analytical power of NRT lies in the ability to analyse chord transitions which are not captured directly as P, L or R transitions, but sequences thereof. For instance, if a transition of one triadic chord *A* to another *B* is captured as the NRO sequence *LP*, then this represents the fact that applying *L* to *A*, then applying *P* to the resulting chord will end with *B*. Lehman (2014) showed that such sequences of P, L and R transitions can model any possible relation between the major and minor triads. Subsequent additions to the theory introduced two inversional operators, *Slide (S)* which exchanges two triads that share a third (such as C major and C♯ Major), and *Nebenverwandt (N)* which transforms a major triad into its minor subdominant, and vice-versa, *(N')* a minor triad into its major dominant (Lehman 2014).

Triadic chromaticism appears in most film scores (Lehman 2014). An example is the soundtrack of the movie *A Beautiful Mind* composed by James Horner. Throughout the whole score, chromatic chord progressions, which can be analysed

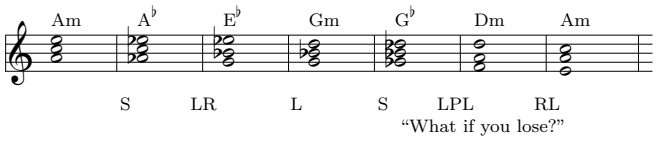

Figure 2: NRT analysis of an excerpt from the soundtrack to the film *Beautiful Mind*. The NRO chord transition sequence is given below the stave.

Table 1: Association of NRO sequences and emotional and/or situational scene elements (Lehman 2014).

NRO Sequence	Emotion/Situation
LP	Antagonism
L	Sorrow, loss
N	Romantic encounters
PRPR	Mortal threats, dangers
RL	Wonderment, success
NRL	Suspense and mystery
RLRL	Heroism (Lydian)
NR	Fantastical
S	Life and death

using NRT, are used to portray the genius of the protagonist. Figure 2 shows a small excerpt of the chord progression used in the cue *"Playing a Game of Go!"*, analysed using NRO. In the film at this point, the music highlights when the lead character loses a game of the board game Go and self doubt creeps in. This is an example of the music supporting an emotional event, which can be captured analytically using NRT. Lehman (2013) links such emotions and events to sequences of NRO operator applications, i.e., sequences of trichord changes. In this way, the mapping of the portrayal of emotional and/or situational aspects of a scene (such as the expression of genius, sorrow and wonderment) and the music accompanying it can be captured analytically. For the set of emotion/situational constructs captured in this way, see table 1. In the excerpt of figure 2, we see that S (life and death) operators, along with L (sorrow, loss) and LP (antagonism) sequences are employed. While the mapping of the music to the emotion/situations in table 1 is not exact, the analysis of the emotional arc of the music matches well the emotions of the film scene being acted out.

Sequences of NRT operators can generate transformations between chords that do not share any notes, and Lehman (2014) states that the transformational complexity and its connection to aural distance can convey feelings and meaning by considering the distance between the original chord and the destination chord. For instance, passages with simple LPR compounds can be associated with relaxation due to the closeness of chords. In contrast, complex combinations can be associated with tension. Furthermore, NRT's transformational approach leads to the possibility of writing tightly voice-lead passages of music. That is, as NROs transform chords into new ones that share common tones, the melodies produced from these operators could present linear progression, making the melodies easy to sing and hence probably more memorable. This is an important feature of soundtracks, which use melodies as story-telling devices (Whittall 2001).

For such film music, NRT is used to analyse chromatic chord progressions that would not be fully justifiable if analysed using a diatonic approach (Lehman 2014) as this type of music makes use of notes foreign to a mode or diatonic scale. Ensuring that the progression stays diatonic requires the use of specific combinations of PLR, therefore not utilizing NRT to its full potential. Furthermore, Lehman states that as NRT operates on pitch classes rather than diatonic notes, NROs disregard enharmonic spelling (e.g. C♯ and D♭ are the same note on a keyboard but the name of the note changes depending on the key signature). In film music analysis and likewise in film music composition, this allows one to focus on other musical features, such as associativity and meaning, that are more important to the genre, while avoiding claims that such chromatic progressions are due to some unjustifiable or irrational music theory.

Table 2: NRT Rewrite rules for major and minor chord starting points. In each case, the starting chord is $\{a,b,c\}$.

NRO	Major	Minor
R	$\{a,b,c+2\}$	$\{a-2,b,c\}$
P	$\{a,b-1,c\}$	$\{a,b+1,c\}$
L	$\{a-1,b,c\}$	$\{a,b,c+1\}$
N	$\{a,b+1,c+1\}$	$\{a-1,b-1,c\}$
N'	$\{a-2,b-2,c\}$	$\{a,b+2,c+2\}$
S	$\{a+1,b,c+1\}$	$\{a-1,b,c-1\}$

Procedural Neo-Riemannian Theory

The transformational relationship of NRT chord progressions is well-suited to work with visual, dialogue and interactive elements of a scene that are meant to evoke feelings. As such, it could be employed in generative systems to write expressive and associative music for films and games. Unfortunately, Cohn's (1998) mathematical description of NRT analytical techniques is frankly overly complicated and confusingly formalised in places. To prepare the theory for implementation in a generative system, we rationalise it here in terms of a set of *conditional rewrite rules*, R, which take a trichord and transform it to another. The conditional check is whether the starting chord is major or minor and each type of chord has six rewrite rules available to transform it, as per table 2. The table prescribes how to transform trichord $\{a,b,c\}$ into a new trichord by adding or subtracting a given number of semitones. As an example, rewrite rule R would take the C-major chord $\{C,E,G\}$ and transform it to $\{C,E,A\}$ (first row, first column of table 2), but would take the C-minor chord $\{C,E\flat,G\}$ and transform it to $\{B\flat,E\flat,G\}$ (first row, second column). Note that for a rewrite rules to be applied, a chord must be permuted into its prime form, i.e., with a being the tonic note, b being a major/minor third above a and c being a fifth above a.

Note that all the NROs rewrite a minor or a major chord to another minor or major chord. Hence, if we start with a major or minor chord, M, we can string together NROs to produce many more transformations of M than in table 2. For instance, starting again with the C-major chord, the sequence LPR would transform it as follows:

$$\{\text{C,E,G}\} \xrightarrow[maj]{\text{L}} \{\text{B,E,G}\} \xrightarrow{\text{I}} \{\text{E,G,B}\} \xrightarrow[min]{\text{P}} \{\text{E,G}\sharp\text{,B}\} \xrightarrow[maj]{\text{R}} \{\text{E,G}\sharp\text{,C}\sharp\}$$

(Note the required step, I, to permute trichord inversion $\{B,E,G\}$ to the prime form $\{E,G,B\}$). This means that a generative system could transform chord $\{C,E,G\}$ into $\{E,G\sharp,C\sharp\}$ in three steps L, P then R but could also transform it directly in one step LPR.

This further means that a sequence of chords can be constructed using the guidelines of table 1 to follow the emotional/situational requirements of a given scene/cut scene in a game or film, or indeed to react live to player actions or other changes in a game. In addition, the length of the sequence could be used as an indication of the strength of emotional/situational change reflected by the chord transformation in the music, as per analytical NRT. This could be done in a stochastic way, perhaps driven by a Hidden Markov Model to introduce variety and surprise in the outputs. Finally, given that entire film soundtracks can be analysed in terms of sequences of NRO rewrite rules, it should be possible to drive the generation of new sequences via machine learning over a corpus of music from a particular composer, or from a particular film.

Related Work

The generation of chord progressions has been investigated for automatic generation of musical harmony, and Wiggins (1999) looks at the notion of intentionality in this respect, which is important for the generation of chord progressions to accompany film/game scenes. Bernardes et. al. (2016) implemented the D'accord harmony generation system which worked over a perceptually motivated tonal interval space. While not using NRT, Monteith et al. (2010) generated music to induce targeted emotions, using statistical techniques such as HMMs, and applied this in (Monteith et al. 2011) to produce affective music to accompany the audio of fairy tales being read.

Chew and Chuan (2011) proposed a style-specific accompaniment system that applies statistical learning to music theory frameworks including NRT. In this work, Neo-Riemannian Theory is used to represent the transitions between adjacent chords and NRT operators (NRO) are used on the Tonnetz, a conceptual lattice diagram representing the tonal space, to build decision trees to statistically learn melody-chord patterns. Given the styles this system mimics (e.g., Rock) and the roman numeral analysis (a type of music analysis where roman numerals are used to represent chords coordinating with scale degrees 1-7) used for the chord progressions, it is clear that this work aimed to create diatonic chord progressions using functional harmony, rather than triadic chromaticism. Similar limitations are seen in other related works such as Amram et al. (2020), where generative chord-based composition is implemented. Here the authors do not implement NRT to its full extent and consider atonality a disadvantage.

Conclusions and Future Work

We have provided a bridge from film and videogame music composition to computational creativity, via a description of the roles of music in films and games, and a procedural reading of an established music analysis technique, namely Neo-Riemannian Theory. The new formalism encapsulates four main points for generative music: (i) a set of rewrite rules for trichords (ii) the sequencing of rewrite rules to provide longer distance chord transforms (iii) a mapping of emotional and/or situational cues to sequences of rewrite rules, and (iv) the observation that chord transform distance is roughly proportional to the strength of the emotional change perceived in the music.

We are building a general-purpose music generation system specifically to aid with film and videogame compositions. At the heart of this will be procedural NRT, and we plan to draw further from music theory when applied to films, to supplement this approach. In particular, we will experiment with the automated invention, repetition and variation of short musical sequences called *leitmotifs*, (Whittall 2001), which can represent characters, emotions, locations and other elements in a film or game. Our aim is to use AI techniques such as HMMs, deep learning, constraint solving and planning to harness NRT and leitmotifs into a system of real utility for composers.

The system will take as input tags that describe emotions, allowing the output to directly follow the scene's emotional arc or, in some less canonical soundtrack examples, provide a contrast to the emotions seen in the media (as suggested by an anonymous reviewer). Producing film and videogame music that deconstructs the viewer's expectations by representing a different emotion than the one seen on screen can create a powerful viewing experience. An example of this technique can be seen in the psychological horror movie Us, where the composer uses a mixture of well-known, happy, upbeat songs to provide a terrifying contrast from the violence that is happening on-screen. The use of unexpected songs which provide a contrasting mood from the visuals can heighten the feelings of fear as they do not offer any hint on what is going to happen, as most canonical soundtracks do. Other times, contrasting moods between music and visuals can be used to provide a comical effect, such as using a feel-good song while a character is going through hardships. We plan to take into account such less-canonical scoring techniques by providing a system that can create soundtracks that respond or contrast the visuals. As aforementioned, this could be achieved by manually annotating the film with mood tags, so that the composer or director can choose how the visuals should be represented, or not represented, by the music in order to create a variety of interesting viewing experiences for the audience.

Future projects include a collaboration with Younès Rabii, developer of the videogame Tea Garden (Pyrofoux 2020). We plan on applying an NRT-based generative system to produce music for games where characters are created dynamically, such as tabletop role-playing games or videogames that feature a component of live automated videogame design (Rabii and Smith Nicholls 2022). Neo-Riemannian Theory would provide an effective framework to produce music that quickly responds to changes in the characters and videogame design, given the possible combinations of Neo-Riemannian operators and the events and emotions they can represent.

We also plan to use our NRT generative system alongside @artbhot (Smith and Colton 2022), a Twitter bot that generates images from user-given text prompts, to create multimedia outputs such as stories with background music. We plan on exploring using music as a background for a story created from a user-given prompt, or creating music from a user-given prompt and using visual imagery to accompany it. As NRT can be used to write music in various genres, it should produce music suitable to represent the story line or the user-given prompt.

Author Contributions

Theoretical work on the rationalisation of NRT for generative purposes was undertaken by the first author, Sara Cardinale, with help from the second author Simon Colton. All authors participated in the writing of this manuscript, with the first author taking the lead role in this.

Acknowledgements

This work has been funded by UKRI and EPSRC as part of the "UKRI Centre for Doctoral Training in Artificial Intelligence and Music", under grant *EP/S022694/1*. We would like to thank the anonymous reviewers for providing helpful and friendly feedback, which has improved this paper.

References

Amram, M.; Fisher, E.; Gul, S.; and Vishne, U. 2020. A transformational modified Markov process for chord-based algorithmic composition. *Mathematical and Computational Applications* 25(3).

Bernardes, G.; Cocharro, D.; Guedes, C.; and Davies, M. 2016. Harmony generation driven by a perceptually motivated tonal interval space. *Computers in Entertainment* 14(2).

Chuan, C.-H., and Chew, E. 2011. Generating and evaluating musical harmonizations that emulate style. *Computer Music Journal* 35:64–82.

Cohn, R. 1998. Introduction to Neo-Riemannian Theory: a survey and a historical perspective. *Journal of Music Theory* 42(2).

Huron, D. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*, volume 1. MIT Press.

Hutchings, P. E., and McCormack, J. 2020. Adaptive music composition for games. *IEEE Transactions on Games* 12(3):270–280.

Lehman, F. 2013. Transformational Analysis and the Representation of Genius in Film Music. *Music Theory Spectrum* 35(1):1–22.

Lehman, F. 2014. Film music and NRT. *Oxford Handbook*.

Lehman, F. 2018. *Hollywood Harmony: Musical Wonder and the Sound of Cinema. Oxford University Press*.

Lewin, D. 1982. A formal theory of generalized tonal functions. *Journal of Music Theory* 26(1):23–60.

Liapis, A.; Yannakakis, G. N.; and Togelius, J. 2014. Computational game creativity. In *Proceedings of the International Conference on Computional Creativity*.

Monteith, K.; Francisco, V.; Martinez, T.; Gervás, P.; and Ventura, D. 2011. Automatic generation of emotionally-targeted soundtracks. In *Proceedings of the International Conference on Computional Creativity*.

Monteith, K.; Martinez, T.; and Ventura, D. 2010. Automatic generation of music for inducing emotive response. In *Proceedings of the International Conference on Computational Creativity*.

Plut, C., and Pasquier, P. 2020. Generative music in video games: State of the art, challenges, and prospects. *Entertainment Computing* 33:100337.

Plut, C. 2017. *The Audience of the Singular (MFA Thesis)*. Simon Fraser University, Vancouver, Canada.

Rabii Younès (Pyrofoux) 2020. Tea Garden. Available at: https://pyrofoux.itch.io/tea-garden.

Rabii Younès (Pyrofoux) and Smith Nicholls Florence. 2022. Choose Your Own Misadventure: AI-Powered Futures for Game Design. Talk given at Game Developers Conference 2022. Available at: https://www.youtube.com/watch?v=qYxRToBWpVo &t=2466sab_channel=Knives%26Paintbrushes.

Rehding, A.; Floud, R.; A,; Johnson, P.; Kallberg, J.; Newcomb, A.; and Solie, R. 2003. *Hugo Riemann and the Birth of Modern Musical Thought*. Cambridge University Press.

Rod, M. 2007. *Music In Video Games, in J. Sexton (ed.) Music, Sound and Multimedia: From the Live to the Virtual*. Edinburgh University Press.

Savery, R., and Weinberg, G. 2020. Shimon the robot film composer and deepscore: An LSTM for generation of film scores based on visual analysis. *arXiv* 2011.07953.

Smith A., and Colton S. 2022. The @artbhot Text-To-Image Twitter Bot. In *Proceedings of the International Conference on Computational Creativity*.

Whalen, Z. 2004. Play along – an approach to videogame music. *The International Journal of Computer Game Research* 4.

Whittall, A. 2001. *Leitmotif*. Oxford University Press.

Wiggins, G. A. 1999. Automated generation of musical harmony: What's missing. In *Proceedings of the International Joint Conference on Artificial Intelligence*.