UNIVERSITY OF SOUTHAMPTON

A Formula for Music Similarity: Utilising music-theoretical approaches in audible perceptions of harmonic similarity

by

Anna Selway

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

in the
Faculty of Humanities
Department of Music

January 2021

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF HUMANITIES

DEPARTMENT OF MUSIC

Doctor of Philosophy

A Formula for Music Similarity: Utilising music-theoretical approaches in audible perceptions of harmonic similarity

by Anna Selway

Harmony appears to have a vital role in listeners' perceptions of musical similarity. However, long-established theories of harmony such as Hugo Riemann's theory of 'harmonic functions' have been under-utilised in the fields of music cognition and perception, and particularly in music information retrieval and forensic musicology. Indeed, it is surprising that such crucial applications still generally rely upon ad-hoc and proprietary methods for determining similarity. My doctoral research explores whether traditional scholarly music-theoretical methods of determining harmony (such as Riemann's theory of harmonic function, and aspects of Schenkerian analysis) could aid in developing better methods for determining similarity. I propose that we would be better able to extract high-level musical features by using traditional music-theoretical methods.

Firstly, I report an initial study that highlights harmonies relevance in participants' classification of audible music similarity. Riemann's theory is then utilised to explain some of the apparent discrepancies in human-annotated harmony datasets; specifically, the Chordify Annotator Subjectivity Dataset, a subset of Chordify's user edit data, and my own annotation study using the song 'Little Bit O' Soul' (Chapters 3, 4, and 6).

This thesis concludes by proposing an adapted version of Riemannian theory (removing the need for a key), which can be applied not only to computationally encoded scores, but also audio and other computationally available data (Chapters 5, and 7). Overall, I show that a Riemannian-based approach that observes the chord labels (not using a score) enables music similarity approaches to explore audible music similarity in more depth. This research not only has significant importance in our understanding of harmonic similarity, but also in understanding how current audio-based extraction methods can incorporate music theory. My use of this theoretical framework in the study of musical similarity could improve methods of determining music similarity used in a variety of other fields, such as the development and implication of copyright law, commercial music sales, music information retrieval extraction and evaluation metrics.

Contents

List of Figures

List of Tables

Research Thesis: Declaration of Authorship

Name: Anna Selway

Title: A Formula for Music Similarity: Utilising music-theoretical approaches in audible perceptions of harmonic similarity

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

   Anna Selway et al., "Explaining Harmonic Inter-Annotator Disagreement using Hugo Riemann's theory of 'Harmonic Function'," *Journal of New Music Research* 49, no. 2 (2020): 136–150.

Signature: .... Date: 12/12/2020

Acknowledgements

I wish to thank all the people whose assistance made the completion of this thesis possible. I would first like to thank the physical and technical contributions of the RCUK Digital Economy Programme, and the Web Science centre for doctoral training. Without their support and funding, this project could not have reached its goal. Thank you for the training and the opportunities to engage internationally with scholars at a variety of conferences, workshops and networking events. I would like to recognise all the individuals who took part in the studies in my thesis. Without your time and commitment, these studies could not have been a success. Alongside this, I would like to thank the volunteers who helped me record 'Little Bit O' Soul' for my final study. Specifically Dan Mar-Molinero for your technical assistance, Dr Andrew Fisher for playing the piano, and Zoe Trivett for singing. I am so thankful for your generosity in time and resources that enabled me to complete this final study.

To my supervisors, Dr David Bretherton, Dr Nicholas Gibbins and Dr Richard Polfreman, thank you all for your time, professional encouragement and guidance. Without your persistence and confidence in this project, its potential would not have been realised. David, as my lead supervisor, I would like to especially thank you for the monumental time, effort, reading and editing you have invested in this project. From this project's first realisation, you have been crucial to my academic prowess and the development of my career.

I would like to express my most profound appreciation for the work of the WiMIR programme, and specifically my mentor Dr Anja Volk. Thank you for your guidance, mentoring and collaboration. The opportunities afforded by you and the team at Utrecht University shaped a significant amount of my thesis. The opportunity to work with datasets collected by yourselves, and data from Chordify made this research possible. I would like to extend my thanks Chordify for providing my fourth chapter's dataset, the breadth of analysis this provided was more considerable than I could have imagined.

I also wish to acknowledge the support and love of my family, especially my husband, Rob, and sister Jess. They kept me going, and this work would not have been possible without their input, including the substantial time throughout the project they both invested in editing and reviewing drafts and the completed work. Finally, I wish to thank and dedicate this thesis to my late grandfather, Richard Garlick. From an early age, he supported and believed in my academic potential.

To Richard Garlick

Part I

Chapter 1

Introduction: Music similarity, and why does it matter?

The notion of musical similarity is an ill-defined and highly subjective concept.[1] However, observing the similarities in music is an intrinsic element of research in multiple disciplines — commercial, legal, and academic. Various definitions of musical similarity form the primary classifications and algorithms used to organise and recommend music by applications such as Apple Music, Spotify, and Last.fm.[2] In the legal domain, forensic musicologists use musical similarity to discover cases of plagiarism; while music theorists use similarity to identify relationships both within and between pieces of music. Each of these applications of musical similarity determine and define what similarity means differently. Two pieces of music can be similar on the basis that the same artist has written them, that they are in the same style or genre, or in terms of more granular similarities created by the musical elements that make up the piece — such as timbre, melody and harmony.[3]

Peter Knees and Markus Shedl (2013) define musical similarity as the distance between the values of a musical feature, either between multiple songs or within the same song.[4]

1. Peter Knees and Markus Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data," *ACM Transactions on Multimedia Computing, Communications, and Applications* 10, no. 1 (2013): 1–21.

2. Jin Ha Lee, "How Similar is too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation," In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 109–114; Adam Berenzweig et al., "A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures," *Computer Music Journal* 28, no. 2 (2004): 63–76.

3. John Stevenson, *Capturing Similarity in Music*, http://www.city.ac.uk/news/2015/june/capturing-similarity-in-music, (Accessed: 18.06.2016), City University London, 2015; Jean-Julien Aucouturier et al., "Music similarity measures: What's the use," in *Proceedings of the 3rd International Society for Music Information Retrieval (ISMIR) Conference* (Paris, 2002), 157–163, ISBN: 2844261663; Lee, "How Similar is too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation"; Kadek Cahya Dewi, Luh Arida, and Ayu Rahning, "Music Recommendation Based on Audio Similarity Using K-Nearest Neighbor," in *Proceedings of the 1st ACIKITA International Conference of Science and Technology (AICST)*, 978 (2011), 124–132; Dmitry Bogdanov et al., "Unifying low-level and high-level music similarity measures," *IEEE Transactions on Multimedia* 13, no. 4 (2011): 687–701.

4. Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data."

Many different musical features can contribute to the definition of similarity, including genre, instrumentation, voice characteristics, melody, harmony, rhythm, and dynamics.[5] As a result, there are multifaceted (commercial) approaches to musical similarity, meaning the definition of similarity for one application may not be the definition for another. Measuring musical similarity is further complicated by the influence of an individual's subjective perception of similarity;[6] for some, two works may be similar, but for others they might not.[7] For example, listeners of classical music are likely to hear differences between the piano works of Bach and Mozart, but an individual who does not listen to classical music could perceive them as similar through their shared genre classification — 'classical music'. Therefore, musical familiarity could be important in terms of how similarity is perceived, as someone with knowledge of a genre may be more adept at perceiving its nuances. Alongside this, higher musical competence could enable a listener to hear similarities between pieces that others may not hear, such as being able to explore and understand the differences between Blues pieces by identifying their differing complex harmonies. For many reasons, subjectively derived similarity measures are inherently problematic, particularly with regards to their capacity to give meaningful results for more 'commercially niche' musical genres such as 'classical music'.

1.1 Similarity in copyright cases

The lack of a single method for judging plagiarism and copyright infringement provides a useful illustration of the complexity of music similarity. The verdict of the Pharell Williams v. Marvin Gaye lawsuit, finding Williams and Thicke guilty of copyright infringement in their song 'Blurred Lines' regarding Gaye's song 'Got to give it up', has prompted significant academic debate.[8] The plaintiff argued that there are striking similarities between the pieces, particularly in terms of instrumentation, timbre, and rhythm — and especially in the use of the salsa rhythm on a cowbell (a percussion instrument).[9] Most criticism has arisen over whether a 'vibe' (timbre, atmosphere, and genre) can be copyrighted;[10] this has lead to much speculation on how this verdict will impact musical influence, and how 'sounding like' another song has led to legal liability.[11]

5. Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data"; Stevenson, *Capturing Similarity in Music*.

6. Malcolm Slaney et al., "Learning a Metric for Music Similarity," In *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference* (Philadelphia), 2008, 313–318.

7. Ibid.

8. Andrea Keifer, "Civil Minutes from Complaint for Infringement, Pharrell Williams et al. vs. Brdigeport Music, Inc., et al.," *United States District Court Central District Of California* Case 2:13 (2014): 1–28, Thicke and Williams appealed the courts' verdict in 2016, but the chord upheld the 2015 verdict in 2018.

9. Ibid.

10. David Post, *Blurred Lines and Copyright Infringement*, Available at: `https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/03/12/blurred-lines-and-copyright-infringement/`, March 2015.

11. Charlotte Tschinder, "Automating Music Similarity Analysis in 'Sound-Alike' Copyright Infringement Cases," *Entertainment, Arts and Sports Law Journal* 25, no. 2 (2014): 60–68.

Charlotte Tschinder highlights that the courts and experts have to analyse a multitude of factors to determine similarity, including 'lyrics, tempo, genre, themes, and specific note sequences'.[12] So much so, that she notes that the experts called upon in these cases filter what is and is not protectable depending on what will provide success for the case they are defending.[13] Also noted by Scott Fruehwald, is the encouraged 'battle of experts' found in this legal setting, through both sides hiring expert witnesses.[14] This means that witnesses aim to succeed in proving (or disproving) the similarity between pieces, rather than providing a conclusive analysis of the pieces across all musical facets.

In direct contrast, the case of Martin Harrington and Thomas Leonard v. Ed Sheeran provides substantial music analytical similarities between the two songs 'Amazing' (2009) and 'Photograph' (2014).[15] This case focused on high-level score features, providing multiple extracts of transcriptions of both songs. The prosecutors noted a 70% overlap between the notes in the two choruses through a note-by-note comparison.[16] The chord progression breakdown highlighted the 'striking' similarities between these two pieces, which the prosecutors claimed made them 'instantly recognisable to the ordinary observer'.[17]

For a verdict of copyright infringement, one has to prove substantial similarities between the two pieces.[18] Though, as highlighted by Jason Palmer (2016), how 'substantial', substantial similarity is, is yet to be pinpointed.[19] As shown in the two examples discussed, the definition of a musical idea or expression, and the musical characteristics to focus on, is inconsistent between litigants.[20] Charles Cronin suggests a formulation for consistency across cases by suggesting music copyright should revert to utilising the 'long-established view of melody, harmony, and rhythm as the *sine quibus non* of a musical work'.[21] In Cronin's opinion, copyright should focus on melody, harmony and rhythm, as these are what collectively make a song unique — according to Cronin, these are the only features important for identifying copyright infringement.

12. Ibid.
13. Ibid.
14. Tschinder, "Automating Music Similarity Analysis in 'Sound-Alike' Copyright Infringement Cases"; E. Scott Fruehwald, "Copyright Infringement of Musical Compositions: A Systematic Appproach," *Akron Law Review* 26, no. 1 (1993): 15–44.
15. Ed Sheeran settled out of court with Harrington and Leonard paying $20 million, along with giving Harrington and Leonard writing credits.
16. Paul H Duvall and Mark L Block, *Complaint for Copyright Infringement, HaloSongs, Inc., Martin Harrington, and Thomas Leonard vs. Edward Christopher Sheeran*, 2016, 1–28.
17. Ibid.
18. Jason Palmer, "'Blurred Lines' Means Changing Focus: Juries Composed of Musical Artists Should Decide Music Copyright Infringement Cases, Not Lay Juries," *Vanderbilt Journal of Entertainment Technology Law* 18, no. 4 (2016): 907–934.
19. Ibid., p.907.
20. Ibid.
21. Charles Cronin, "Seeing Is Believing: The Ongoing Significance of Symbolic Representations of Musical Works in Copyright Infringement Disputes," *Colorado Technology Law Journal* 16, no. 2 (2018): p.228.

1.2 Similarity in music recommendation

The rapid developments in digital distribution, enabled by the technological advancements in the encoding and compression of audio signals (such as the introduction of the MP3 standard), has changed the way we 'use' music in our daily lives.[22] Portable music players and the World Wide Web have enabled music to accompany many daily activities. Until recently, the way music was searched for and discovered remained much the same: with songs forming collections in albums promoted via the radio or interpersonal recommendation. The way we recommend music has also changed as a result of these developments in distribution, with music recommender systems providing information filtering, to assist users in discovering new music.[23] Music recommendation utilises listening behaviours, collaborative filtering, meta-data and content-based feature extraction, each facet on their own or in collaboration with one another, to suggest music to a user based on their preferences.

Collaborative filtering is perhaps the most prominent method of music recommendation. It works by filtering items based on the opinions of other people 'like you'.[24] Similarity, in this approach, is determined by measuring the 'distance' between users, such as using a Pearson correlation coefficient and clustering users based on their clickstream data.[25] Although collaborative filtering is essential for recommending music, this thesis critiques current approaches to musical similarity. For this purpose, the appropriate recommender methods (those which focus on the musical similarity of the pieces) include metadata and, most relevant, content-based feature extraction.

Metadata is 'data about data'; in this case, contextual data about the track of music — that is not the track itself, such as the artist, date of release, genre, style and copyright owner.[26] In 2008 the most popular way of measuring the similarity of two pieces of music

22. Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data."

23. Yading Song, Simon Dixon, and Marcus Pearce, "A Survey of Music Recommendation Systems and Future Perspectives," in *9th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, June (London, 2012), 19–22; Michael .A. Casey et al., "Content-Based Music Information Retrieval: Current Directions and Future Challenges," 96, no. 4 (2008): 668–696.

24. J. Ben Schafer et al., "Collaborative Filtering Recommender Systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*, ed. Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl (Berlin: Springer-Verlag, 2007), 291–324; Billy Yapriady and Alexandra Uitdenbogerd, "Combining Demographic Data with Collaborative Filtering for Automatic Music Recommendation," In *Proceedings of the 9th International Conference of Knowledge-Based Intelligent Information and Engineering Systems* (Australia), 2005, 201–207.

25. Bamshad Mobasher et al., "Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data," in *Workshop Intelligent techniques for Web Personalization*, ed. Sarabjot Singh Anand and Bamshad Mobasher (Berlin: S, 2001), 53–60; U Shardanad and P Maes, "Social Information Filtering: Algorithms for Automating Word of Mouth.," In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence* (Madison), 1995, John Breese, David Heckerman, and Carl Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," In *Proceedings of the ACM Conference on Computer Human Interaction* (Denver), 1998, 210–217.

26. Tom Turner, "What is Metadata?," *Kaleidoscope* 10, no. 7 (2002): 1–3, doi:10.1007/978-1-4020-6869-0_2; Lee, "How Similar is too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation."

was still by comparing their contextual metadata.[27] The music and podcast platform Pandora has commercially utilised metadata to recommend music, through their Music Genome Project.[28] This project is the foundation of their system, which used trained musicologists to study and collect tags for the songs in their collection, and Pandora boasts a high inter-reviewer agreement, even stating this makes them 'musically objective'.[29] These 150–500 tags per song are mostly genre-based including tags such as 'Afro-Latin roots' and 'Electric Piano Riffs'.[30]

The most frequently used metadata is genre, which ranks as a 'strong' or 'successful measure of similarity between songs'.[31] However, Jason Neal (2014) highlights that genre categorisation can be less objective in a variety of manners, including '(1) the number of genres, and (2) the granularity to which levels of subgenres are parsed out';[32] although Neal does observe that songs within a genre, no matter how the genres are defined, will have 'common musical facets/traits'.[33] Although metadata approaches have afforded a level of similarity between pieces, generally, limitations exist in their ability to enable cross-genre comparisons.[34]

Genre-based recommendation requires the categorisation of an artist, or a piece of music, into a single genre. However, a single piece can sometimes be categorised according to several genres, even moving genre as their music changes, or as the artist gains more publicity.[35] For example, on Apple Music, the music of Lil Nas X was initially categorised as 'New Mexico Music', or 'Country'. However, as his music gained popularity (with his song 'Old Town Road'), his music has been moved to the 'Alternative' genre.[36] The 'Alternative' genre is a collection of music post-1970 that falls outside of a conventional genre (for example, another 'Alternative' artist is Nirvana).[37] Though what is interesting, is musicologists would often categorise the music of Nirvana as grunge. The 'Alternative' genre seems to have an ever changing meaning being a very generic

27. Casey et al., "Content-Based Music Information Retrieval: Current Directions and Future Challenges."
28. Pandora Media, 'About the Music Genome Project', Pandora, 2020, https://www.pandora.com/about/mgp [accessed 1st March 2020].
29. Casey et al., "Content-Based Music Information Retrieval: Current Directions and Future Challenges"; Douglas Turnbull, Luke Barrington, and Gert R. Lanckriet, "Five Approaches to Collecting Tags for Music," In *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference* (Philadelphia), 2008, 225–230.
30. Turnbull, Barrington, and Lanckriet, "Five Approaches to Collecting Tags for Music."
31. Beth Logan and Ariel Salomon, "A Music Similarity Function based on Signal Analysis," *IEEE International Conference on Multimedia and Expo (ICME)* (Tokyo), 2001, 745–748; Aucouturier et al., "Music similarity measures: What's the use."
32. Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data"; Jason Neal, *Defining Musical Similarity: Genre and Beyond*, http://www.asis.org/SocialMedia/?p=66, (Accessed: 25.06.2016), The Information Associaton for the Information Age (ASIS&T), 2014.
33. Neal, *Defining Musical Similarity: Genre and Beyond*.
34. Ibid.
35. Mario J. Lucero, *Music Streaming Services Mishandle our Data—and our Culture is Paying for it*, Available at: https://qz.com/1773480/the-problem-with-how-the-music-streaming-industry-handles-data/, January 2020.
36. Ibid.
37. Ibid.

category that varies between music platforms. This means that music within this genre does not feature 'common musical facets/traits', as previously stated by Neal (2014) as a success of genre similarity between music.[38] Therefore, a song or artist can straddle the borders of multiple genres, be categorised differently between systems, and the music could be considered similar to a number of other genres.

The capacity for music to accompany our everyday life has led to people seeking music that accompanies a 'mood' or 'event', instead of according to genre. This has seen a considerable growth in mood-based recommender work featuring at conferences such as the International Society for Music Information Retrieval's Evaluation eXchange (MIREX) competition's category 'Audio Mood Classification task'.[39] The mainstream media has often coined the term 'the death of the music genre' to refer to the fact that we now ask Alexa to play 'happy music' instead of 'popular music'.[40] This thesis is most interested in content-based recommendation as a clear way of defining the similarity of two pieces of music based on their content, not just the category in which they fall. Content-based recommendation does not consider the genre, category, or social tags that the piece is classified within; instead, it looks at what makes the pieces musically similar.

Feature	**Content**
Timbre	Quality and type of produced sound
Melody	Sequence of notes in a desired rhythm, perceived as a single entity
Bass Line	The lowest sequence of notes in a rhythmic pattern
Rhythm	Patterns of sound onsets
Pitch	The quality of sound of one given note
Harmony	Chord progression
Key	The adherence of notes and harmonies to key profiles
Structure	The layout of the composition — as divided into sections
Tempo	The speed of the piece of music

TABLE 1.1: Musical features commonly used in content-based similarity comparisons, referring to the elements of the music which would need to be analysed to extract each feature, based on the table by Orio, Nicola. 'Music Retrieval: A Tutorial and Review.' *Foundation and Trends in Information Retrieval* (Hanover) 1, 1 (2006): 1–96.

Content-based recommendation utilises musical features in searching for and recommending music (see Table 1.1). Content-based extraction is one of the central tasks of the Music Information Retrieval (MIR) community.[41] Most content-based MIR systems use some form of audio (or, less prominently, a symbolic format) to extract raw data (such as scalars, vectors or matrixes) and convert them into a musical feature perceived

38. Lucero, *Music Streaming Services Mishandle our Data—and our Culture is Paying for it*.
39. https://www.music-ir.org/mirex/wiki/2019:Audio_K-POP_Mood_Classification
40. Michael. Whalen, *The Death of the Music Genre (mostly. . .)*, Available at: https://medium.com/@michaeljwhalen/the-death-of-the-music-genre-mostly-7039094302f, April 2018.
41. Bogdanov et al., "Unifying low-level and high-level music similarity measures"; J. Stephen Downie, "The Music Information Retrieval Evaluation Exchange (2005-2007): A Window into Music Information Retrieval Research," *Acoustical Science and Technology* 29, no. 4 (2008): 247–255; Alicja A. Wieczoekowska, "Music Information Retrieval," chap. 216 in *Encyclopedia of Data Warehousing and Mining*, 2nd, ed. John Wang (Hershey: IGI Global, 2009), 1396–1402.

by the human ear (such as those in Table 1.1).[42] Comparing these features enables us to obtain the distance between two songs.[43] Music is highly complex, so there is no specific combination of musical features said to provide the 'best' similarity comparison. However, the features most frequently utilised are: tempo, timbre, melody, rhythm, and harmony.[44] Computationally, the most easy to extract are low-level features, which are those 'mechanically recovered' from the audio or score, such as the amplitude, the tempo, and the timbre.[45] In contrast, high-level features combine low-level features with content; these include the key, pitch, tempo, notes, and phrases. Currently their has been a prominence in the identification of a pieces mood in the MIR community. I have not included mood in my list of features, as mood arguably is created through the co-existence of a variety of the features present in Table 1.1, including harmony, lyrics and rhythm,[46] along with non-content based information such as social tags.[47]

Musical features are essential in the perception of similarity.[48] Research conducted by Jin Ha Lee (2011) explored the potential use of audio music similarity retrieval systems for generating playlists based on pieces of music deemed similar.[49] Lee found that participants perceived similarity not just through metadata similarities, but also looked for features such as tempo and instrumentation. Common responses to the similarity of the pieces included 'the songs all have pretty much the same tempo'.[50] Recent techniques in music similarity have focused on creating musical fingerprints to compare many musical features of a song together.[51] Tools like Shazam have used fingerprinting; for each song, they create a compact, unique feature representation and try to match a query song with equivalent 'fingerprints'.[52]

Due to the complexity of high-level tasks in MIR, the developments of feature extraction to be used in content-based similarity have been limited, suffering from issues of accuracy

42. Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data."
43. Bogdanov et al., "Unifying low-level and high-level music similarity measures"; Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data."
44. Stevenson, *Capturing Similarity in Music*; Aucouturier et al., "Music similarity measures: What's the use"; Lee, "How Similar is too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation"; Dewi, Arida, and Rahning, "Music Recommendation Based on Audio Similarity Using K-Nearest Neighbor"; Bogdanov et al., "Unifying low-level and high-level music similarity measures."
45. Casey et al., "Content-Based Music Information Retrieval: Current Directions and Future Challenges."
46. Xiao Hu and Stephen J. Downie, In *Proceedings of the 11th International Society for Music Information Retrieval (ISMIR) Conference* (Utrecht), 2010, 619–624; Matt McVicar, Tim Freeman, and Tijl De Bie, "Mining the Correlation Between Lyrical and Audio Features and the Emergence of Mood," in In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami, 2011), 783–788.
47. Cyril Laurier et al., "Music Mood Representations from Social Tags," In *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference* (Utrecht), 2009, 381–386.
48. Lee, "How Similar is too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation."
49. Ibid.
50. Ibid.
51. Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data."
52. Ibid.

with high-level tasks such as melody and key extraction.[53] These limitations in high-level tasks have created a 'glass ceiling effect' on the future development of content-based similarity tasks. In terms of the musicianship required, the features often termed 'high-level' by the MIR community are relatively basic, matching the expected skill level of an A-level (or high-standard GCSE) student.[54] This suggests that there is higher musicianship available to be utilised within content-based similarity tasks.

The field of music theory could provide a solution to the current limitations of MIR in extracting high-level musical features. MIR often overlooks music-theoretical approaches as they frequently require a musical score. However, the advancements in music encoding have enabled a computational musical score to be available, driven by the academic-led Music Encoding Initiative (MEI), the publisher-led MusicXML, and other formats that are no longer as prominent, such as HUMDRUM and LilyPond. Encoding in music is the process of converting the visual representation of music (the score) into a series of computational characters (i.e. Extensible Markup Language (XML)), which can be searched and retrieved computationally.[55] These developments in the encoding of sheet music have enabled computational score-based analysis, in turn enabling their use in MIR techniques.[56]

This thesis explores the use of traditional music-theoretical approaches in music similarity. In this chapter, I will provide a literature review of the current prominent music theories, and their application to musical similarity (including Schenkerian analysis, *Formenlehre*, Riemannian, and neo-Riemannian theory). Following this, a review of the existing computational approaches to music theory will provide an insight into where research is lacking in this domain. The different music-theoretical approaches to similarity, examined further in Chapter 2, provide a rationale for the focus of this thesis on the potential of Riemannian theory in computational musical similarity. This thesis hypothesises that if one annotator can perceive a different chord to another annotator, then the two chords could be seen as perceptually similar, as they are audibly mistakable.

53. Casey et al., "Content-Based Music Information Retrieval: Current Directions and Future Challenges."

54. 'GCSE', or the General Certificate of Secondary Education, is the examination typically taken by students at the end of their 11th year of compulsory full-time school education, typically aged 15–16. 'AS', or advanced subsidiary level, is the first component of the A-Level (Advanced Level) qualification in optional further education, typically taken around age 16–17. With the complete A-Level constructed of both this AS and second-year examinations typically sat around age 17–18. A good A-Level in Music is often required for a student to study music in higher education (such as at university). These are the examinations sat in the English, Welsh and Northern Irish education system.

55. John Daintith and Edmund Wright, *Code*, Available at: http://www.oxfordreference.com/view/10.1093/acref/9780199234004.001.0001/acref-9780199234004, 2008.

56. Wijnard Schepens and Marc Leman, "Chronicle: XML-Representation of Symbolic Music and Other Complex Time Structures," in *Structuring Music through Markup Language: Designs and Architectures*, ed. Jacques Steyn (Hershey: Information Science Reference, 2012), 99–118; Gerant Wiggins, "Computer Representation of Music in the Research Environment," in *Modern Methods for Musicology: Prospects, Proposals, and Realities*, ed. Tim Crawford (Farnham: Ashgate, 2009), 7–27; Donald Byrd and Eric Isacson, "A Music Representation Requirement Specification for Academic," *Computer Music Journal* 27, no. 4 (2003): 43–57; Jacques Steyn, "Framework for a Music Markup Language," *Proceedings of the 1st International Conference of Musical Applications using XML (MAX)*, 2002, 22–29.

1.3 Music-theoretical approaches to similarity

Music analysis is the process of understanding or 'getting to grips' with a piece of music, in order to discover or realise how the piece works.[57] By utilising a set of methodologies (music theories), we perform a 'close reading' of an individual or multiple pieces to describe, rather than judge, the piece.[58] In general, these methodologies take the form of either labelling specific objects within the score, or providing a structural analysis of the piece. Labelling approaches focus on a single specific musical feature in isolation.[59] Interestingly, these are all high-level features that one derives from the musical score to describe the piece.[60] The musical feature of harmony is associated with a variety of labelling approaches, most of which have grown from the study and teaching of harmony — *Harmonielehre*, famously disliked by Arnold Schoenberg.[61] One of the most prominent and influential of these theories is Riemannian theory, which looks to label chords based on their functional role, and their relationship to the tonic, dominant and subdominant.[62] Neo-Riemannian theory provides a similar labelling style analysis, but looks at the relationship of chords on a horizontal plane to observe how they transform or move from one another across the length of a piece.[63] Similarly, this theory is interested purely in the feature of harmony, specifically the understanding of chromatic music that is not tonally unified.[64] *Formenlehre*, also a labelling theory, uses both harmonic and melodic information in its labelling.[65] Often, an analyst will select one musical facet for their analysis, leading to very different categorisations of a single piece's form. A famous example of this disagreement in classification is sonata form — whether it is in two-part form (harmonic) or three-part form (thematic).

In contrast, structural analysis focuses on the relationships between different components in a piece.[66] As described by Philip Kirlin and Paul Utgoff (2008), 'structure' in

57. Ian D. Bent and Anthony Pople, "Analysis," in *Grove Music Online*, ed. Stanley Sadie (Oxford University Press, 2001); Nicholas Cook, *A Guide to Musical Analysis* (London: Butler / Tanner Ltd., 1987).

58. Bent and Pople, "Analysis."

59. Phillip B Kirlin and Paul E. Utgoff, "A Framework for Automated Schenkerian Analysis," In *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference* (Philadelphia), 2008, 363–368; Alan Marsden, "Generative Structural Representation of Tonal Music.," *Journal of New Music Research* 34, no. 4 (2005): 409–428.

60. Marsden, "Generative Structural Representation of Tonal Music."

61. Carl Dahlhaus et al., "Harmony," in *Oxford Music Online* (Oxford University Press, 2009); O.W Neighbour, "Schoenberg [Schönberg], Arnold," in *Oxford Music Online* (Oxford University Press, 2001).

62. Brian Hyer, "Tonality," in *The Cambridge History of Western Music Theory*, 3rd, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2006), 726–752; Brian Hyer, "What is a Function?," in *The Oxford Handbook of neo-Riemannian Music Theories*, ed. Edward Gollin and Alexander Rehding (New York: Oxford University Press, 2011), 92–139.

63. Richard Cohn, "Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective.," *Journal of Music Theory* 42, no. 2 (1998): 167–180.

64. Ibid.

65. William E. Caplin, "What are Formal Functions?," in *Music Form, Forms, Formenlehre: Three Methodological Reflections*, 2nd, ed. Pieter Bergé (Leuven: Leuven University Press, 2010), 21–40; Dahlhaus et al., "Harmony"; Charles J. Smith, "Musical Form and Fundamental Structure: An Investigation of Schenker's 'Formenlehre'," *Music Analysis* 15, nos. 2/3 (1996): 191–297.

66. Kirlin and Utgoff, "A Framework for Automated Schenkerian Analysis."

this context refers to 'the complete fabric of the composition as established by melody, counterpoint, and harmony in combination'.[67] These analytical methods do not observe a single musical feature, but instead focus on their relationships and how they collectively form the piece. The most famous theoretical approach in this method is Schenkerian analysis, which looks at the relationship between melody, harmony, counterpoint, and form. Heinrich Schenker's theory observes how these components come together to shape a fundamental structure (Ursatz) into the foreground music we hear as the piece.[68]

1.3.1 Riemannian Theory

Hugo Riemann's discourse on harmony moves away from the traditional emphasis on a triad's relationship to the tonic, instead focusing on what Riemann coined as a functional theory of harmony.[69] This approach emphasises the harmonic purpose of a chord instead of its identity. Riemann insisted that he found harmonic functions from harmonic dualism; that a substitution was a way of inverting a chord between its major and minor counterpart.[70] According to Riemann, there are three types of harmonic function: the tonic (`T`), dominant (`D`), and subdominant (`S`).[71] These functions do not refer to a single chord, as is traditional in Western music-theory (i.e. the tonic being scale degree `I`, the dominant scale degree `V`, and subdominant scale degree `IV`). Instead, Riemann shows how many chords can assert the same harmonic function as others through utilising a set of substitutions.[72]

Riemann identifies that a piece of music features a continual series of cadential sequences, in the functional structure of `T`–`S`–`D`–`T`. In its most basic form, this would be the chord progression `I`–`IV`–`V`–`I`, but Riemann details a set of three core harmonic substitutions that enable other chords to provide the tonic, dominant and subdominant functions of this progression — the *Variante*, *Parallele* and *Leittonwechsel*.[73] Using these core substitutions, the chords built on the first (`I`), fourth (`IV`) and fifth (`V`) scale degrees change harmonically while still retaining the chord's function. The first, the *Variante*, substitutes the chord with its opposing mode. For example, in Figure 1.1 (a) C major

67. Kirlin and Utgoff, "A Framework for Automated Schenkerian Analysis."
68. Allen Forte and Steven E. Gilbert, *Introduction to Schenkerian Analysis* (New York: W. W. Norton & Company, Inc., 1982).
69. Hugo Riemann, *Harmony Simplified: or, the Theory of the Tonal Functions of Chords* (London: Augener, 1896); Hugo Riemann, "Ideas for a Study 'On the Imagination of Tone'," trans. Robert W. Wason and Elizabeth West Marvin, *Journal of Music Theory* 36, no. 1 (1992): 81–117.
70. David W. Bernstein, "Nineteenth Century Harmonic Theory," in *The Cambridge History of Western Music Theory*, 3rd, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2006), 778–817; Robert W. Wason, "Music Practica: Music Theory as Pedgagogy," in *The Cambridge History of Western Music Theory*, 3rd, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2006), 46–77.
71. Hyer, "What is a Function?"
72. Daniel Harrison, *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of its Precedents* (Chicago: The University of Chicago Press, 1994).
73. Harrison, *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of its Precedents*; Hyer, "Tonality"; Hyer, "What is a Function?"

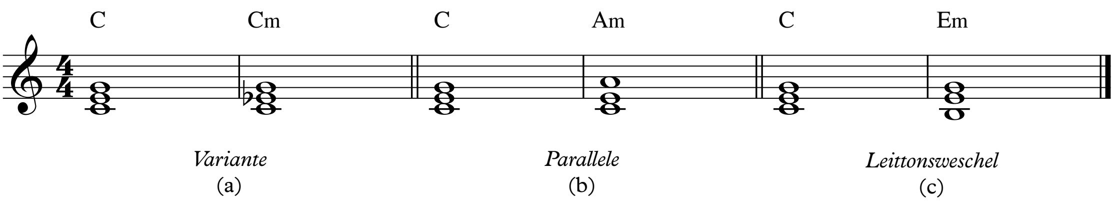


FIGURE 1.1: The three basic Riemannian substitutions, *Variante*, *Parallele* and *Leittonwechsel*.

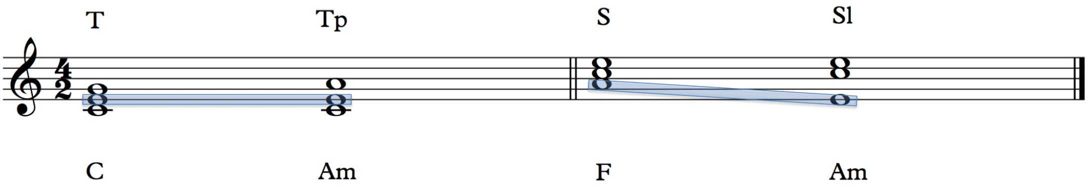


FIGURE 1.2: The chord A minor (`vi`), in the key of C major. According to Riemannian theory this can be either `Tp` through a *Parallele* substitution of the tonic (C major), or `Sl` through a *Leittonwechsel* substitution of the subdominant (F major).

(`T`) and C minor (`t`) using the *Variante* substitution — this moves the third up or down a semitone (down for major to minor, and up for minor to major).

The second substitution defined is the *Parallele*, commonly known in English harmony discourse as the 'relative', which connects the major and minor triads whose roots are a third apart. For example, in Figure 1.1 (b), substituting the chord C major (`T`) with A minor (`Tp`) by moving the fifth (G) of C major up a tone to create the root of A minor. In reverse, substituting the chord A minor (`t`) for C major (`tP`) requires the root of A minor to move down a tone to G. The final core substitution defined is the *Leittonwechsel*. This substitution establishes a relationship between the major and minor triads whose roots are a major third apart (e.g. C major and E minor). For this substitution, a different pitch-class moves depending on whether the chord it is substituting is major or minor. The root of a major chord (C) moves down a semitone (B) to reach the minor chord, whereas a minor chord's fifth (B) moves up a semitone (C) to create the major substitution (Figure 1.1 (c)).

Using these approaches to substitution, each chord of a major or minor scale can hold at least one possible function, though some chords, such as `iii` and `vi`, can hold multiple functions (see Figure 1.2 for the example of the chord A minor in the key of C major holding both the tonic and subdominant function). Table 1.2 shows a definitive list of all the chords that can hold each function.

Function	**Chord (substitution)**
Tonic	`I (T), iii (Tl), vi (Tp), i (t), ♭III (tP), ♭VI (tL)`
Subdominant	`IV (S), ii (Sp), vi (Sl), iv (s), ♭II (sL), ♭VI (sP)`
Dominant	`V (D), iii (Dp), vii (Dl), v (d), ♭III (dL), ♭VII (dP)`

Table 1.2: The different chords that can hold each one of the three harmonic functions (tonic, dominant and subdominant), through different substitutions.

Riemannian theory enables us to see chords otherwise assumed to be different, 'dissimilar' chordal structures (by purely harmonic composition) as similar — for example, chords `I` and `iii`. This means of determining the similarity between chords related by a substitution has been prominent in a number of copyright infringement cases — for example, in the case of Martin Harrington and Thomas Leonard v. Ed Sheeran discussed in Section 1.1. Specifically, the court documents highlight that the only harmonic difference between the two choruses feature 'in measure 12 where "Amazing" uses Gm7, (Gm7 = G minor with addition seventh) whereas "Photograph" uses B♭'.[74] The plaintiff goes on to discuss how these chords are related because B♭ major functions as the relative major of Gm7. To make this argument, the plaintiffs transposed 'Photograph' into the same key as 'Amazing' (E♭ major). They compared the chords for their similarities in function, demonstrating that in bar 12 the B♭ major chord used by Sheeran can be seen as merely a substitution of the chord Gm7 (using the *Parallele* substitution).[75] Without knowing it, Harrington and Leonard's case for similarity relied on the innate similarity perceived in western tonal harmony, codified by Riemann, between a chord and its *Parallele* substitution. The music perception research of Carol Krumhansl, Jamshed Bharucha and Edward Kessler (1982) also highlights this perceptual relatedness between chords related by a *Parallele* substitution; they found a high relatedness between a major key and its relative minor.[76] Also interesting is the plaintiffs transposition of the key of 'Photograph', effectively saying that the chords perform the same function within their original key as each other, and when transposed the functional link between them is undeniable.

74. Duvall and Block, *Complaint for Copyright Infringement, HaloSongs, Inc., Martin Harrington, and Thomas Leonard vs. Edward Christopher Sheeran*, p.18.

75. Johanna Devaney et al., "Theme and Variation Encodings with Roman Numerals (TAVERN): A New Data Set for Symbolic Music Analysis," In *Proceedings of the 16th annual International Society for Music Information Retrieval (ISMIR) Conference* (Málaga), 2015, 728–734.

76. Carol L. Krumhansl, Jamshed J. Bharucha, and Edward J. Kessler, "Perceived Harmonic Structure of Chords in Three Related Musical Keys," *Journal of Experimental Psychology: Human Perception and Performance* 8, no. 1 (1982): 24–36.

1.3.2 Neo-Riemannian Theory

Sometimes called 'transformational theory', neo-Riemannian theory is the term used to describe the examination of Riemann's ideas for the study of chromatic music.[77] Prominent analysts involved in this research include David Lewin, Brian Hyer, Richard Cohn, and Henry Klumpenhouwer. By stripping the concepts of tonal centricity (focusing on the chords' relation in a key) and dualism from Riemannian theory, neo-Riemannian theory utilises Riemann's concepts of voice-leading and observes the relationship between triads using transformations.[78] At the forefront of this theory is David Lewin, who looked to view the relationship between chords based on the pitch-classes they share. Lewin utilised Riemann's table of tonal relations (the *Tonnetz*) more prominently to visualise a piece of music as a series of harmonic transformations.[79] Transformations introduce the assumption of enharmonic equivalence (an aspect Riemann did not adopt), and re-conceptualises Riemann's theory of substitutions instead perceiving the relationship between triads as they move horizontally through a piece — rather than substituting one chord with another of the same function to fit a cadential formula.[80]

The three core transformations adapted from Riemann's substitutions, according to Cohn, are P, R and L — Parallel, Relative and Leading-tone exchange.[81] Similarly to Riemann's substitutions, each of these transformations alters the mode of a triad, while changing only a single tone.[82] The P transformation is similar to Riemann's *Variante* substitution; this looks to see a chord change to its parallel major or minor triad (moving the third of the triad up or down a semitone, bars 1–2 of Figure 1.3). The R transformation is similar to Riemann's *Parallele* substitution (note that for Riemannian and neo-Riemannian theory, the *Parallele* substitution and P transformation bear no relation, due to terminology differences between German and English). This transformation moves a triad to its relative major or minor key (see bars 3–4 of Figure 1.3), moving the fifth of the triad up (major-minor) or down (minor-major) a tone. The final core transformation, L, is similar to Riemann's *Leittonwechsel* substitution, moving the root

77. Richard Cohn et al., *Harmony*, Available at https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000050818, 2001.

78. Ibid.

79. Richard Cohn, "Maximally Smooth Cycles, Hexatonic Systems, and the Analysis of Late-Romantic Triadic Progressions.," *Music Analysis* 15, no. 1 (1996): 9–40; Richard Cohn, *Audacious Euphony: Chromaticism and the Triad's Second Nature.* (New York: Oxford University Press., 2012); Richard Cohn, "As Wonderful as Star Clusters: Instruments for Gazing at Tonality in Schubert," *19th Century Music* 22, no. 3 (1999): 213–232; Nora Engebretsen, "Neo-Riemannian Perspectives on the Harmonieschritte, with a Translation of Riemann's Systematik der Harmonieschritte.," in *The Oxford Handbook of neo-Riemannian Music Theories*, ed. Edward Gollin and Alexander Rehding (New York: Oxford University Press, 2011), 351–381. The *Tonnetz* was not a new idea conceived by Riemann, but developed in the work of Leonhard Euler (1739).

80. Cohn, "Maximally Smooth Cycles, Hexatonic Systems, and the Analysis of Late-Romantic Triadic Progressions."

81. Engebretsen, "Neo-Riemannian Perspectives on the Harmonieschritte, with a Translation of Riemann's Systematik der Harmonieschritte."

82. Carol L. Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations," *Journal of Music Theory* 42, no. 2 (1998): 256–281.

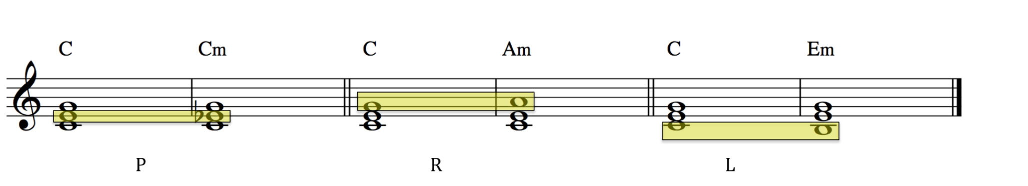


FIGURE 1.3: The three Neo-Riemannian transformations, P, L and R on the C major triad, with the transformation and transformed tones highlighted.

of a major chord down a semitone, or the fifth of a minor chord up a semitone. The L transformation converts a triad to its median (see bars 5–6 of Figure 1.3).

Other transformations exist alongside these core transformations, including the D 'dominant' transformation, N 'nebenverwandt' transformation, and S 'slide' transformation.[83] These transformations do not hold the same status as the core transformations, due to their equivalence to a combination of the P, L and R transformations. For example, the D transformation is equivalent to a combination of L and R transformations. Applying the D transformation to C major gives us G major (the chord's dominant); this is the same as applying the L transformation to C major (to reach E minor), and then the R transformation (to get to G major) — therefore applying LR to C major gives us G major. The use of combined transformations is another difference between Riemann's theory and neo-Riemannian theory, since by using this concatenation of transformations it is possible to use a sequence of P, L, and R to map any chord onto another.[84] Specific combinations of transformations that form repeating patterns are called harmonic 'cycles'; specific named examples include the hexatonic cycle (PLPLPL) and the octatonic cycle (PRPRPRPR). Cohn found that these cycles, using single voice changing progressions, led to 'parsimonious voice-leading'.[85] He argued that voice-leading is naturally parsimonious when considering two chords linked by a single core transformation (L, P, R).[86] From this, Cohn designed the hexatonic system, a set of four maximally smooth cycles, where each chord in the cycle is linked to the next by sharing two common tones.[87] To create these maximally smooth cycles, the 24 major and minor triads are split across four 'hemispheres' (north, south, east, west) where they each form a

83. Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations"; Cohn, *Audacious Euphony: Chromaticism and the Triad's Second Nature.*; Richard Cohn, "Neo-Riemannian Operations, Parsimonious Trichords and their "Tonnetz" representations.," *Journal of Music Theory* 41, no. 1 (1997): 1–66.

84. Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations"; Cohn, *Audacious Euphony: Chromaticism and the Triad's Second Nature.*; Cohn, "Neo-Riemannian Operations, Parsimonious Trichords and their "Tonnetz" representations."

85. Cohn, "Neo-Riemannian Operations, Parsimonious Trichords and their "Tonnetz" representations."

86. Ibid.

87. Cohn, "As Wonderful as Star Clusters: Instruments for Gazing at Tonality in Schubert."

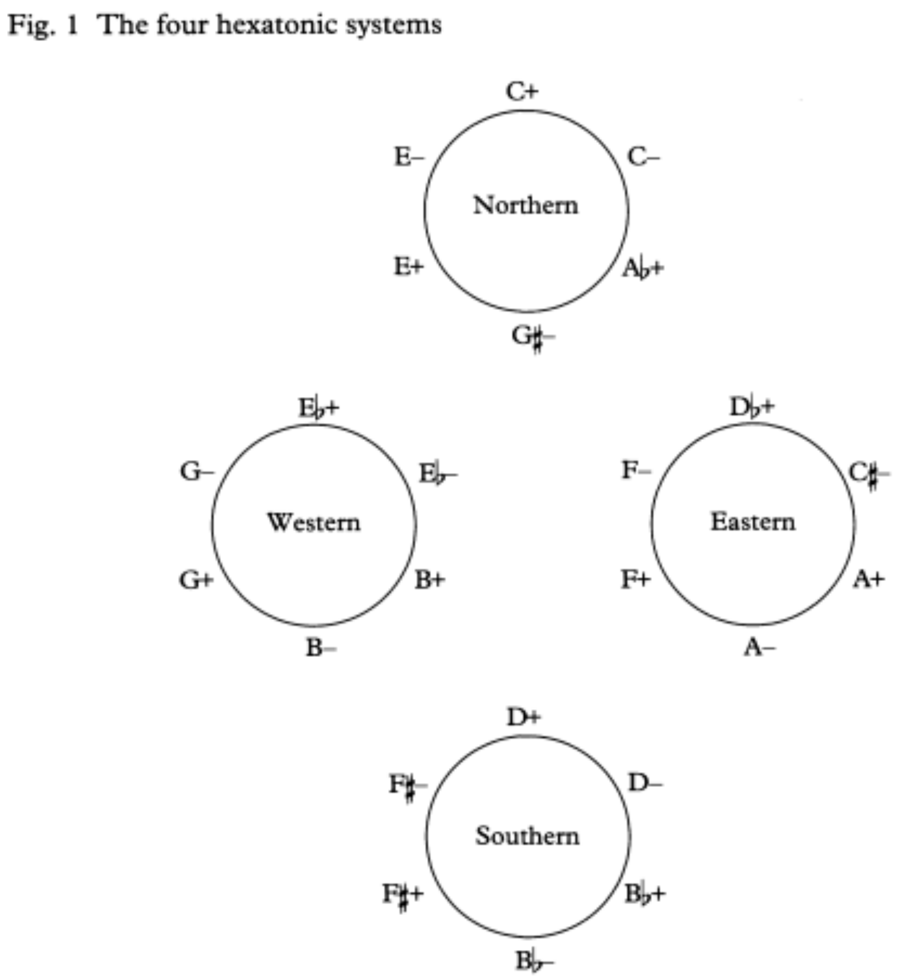


FIGURE 1.4: Richard Cohn's four hemisphere's, Figure 1 as detailed in Cohn, Richard. 'As Wonderful as Star Clusters: Instruments for Gazing at Tonality in Schubert'. *19th Century Music* 22, 3 (1999): 213–232.

hexatonic cycle within their sphere (PLPLPL) — Figure 1.4 details the chords within each hemisphere.

The 'table of tonal relations' or the *Tonnetz* is another graphical structure important to neo-Riemannian theory. Utilised in Riemann's theory as well, this graph connects three independent axes: one with intervals of a perfect fifth (horizontal), one with intervals of a major third and the final with intervals of a minor third (Figure 1.5).[88] Each triangular region, often a triad (though this could represent tones or keys), can be transformed

88. Cohn, "Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective."; Edward Gollin, "Some Aspects of Three-Dimensional 'Tonnetz'.," *Journal of Music Theory* 42, no. 2 (1998): 195–206.

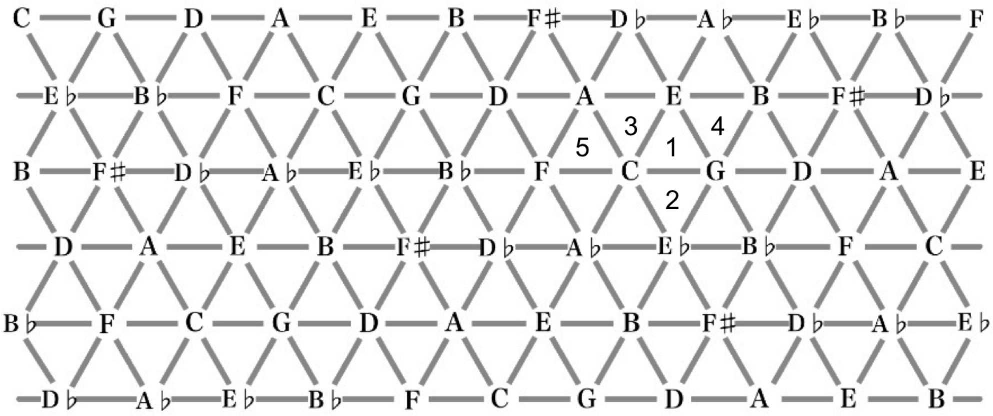


Figure 1.5: The Neo-Riemannian *Tonnetz* graph. With labelled triads for use in discussing the L, P and R transformations, and the shortest path concept that was detailed by Andrew J. Milne and Simon Holland, 'Empirically testing *Tonnetz*, voice-leading and spectral models of perceived triadic distance'., *Journal of Mathematics and Music* 10, 1 (2016): 59–85.

into another through inverting around different edges to reach adjacent triangular regions (a single flip equates to a single transformation, which retains parsimonious voice leading).[89] Triads that are related by a P transformation share a horizontal edge (for example, chords 1 and 2 in Figure 1.5). Triads related by an R transformation share a right-leaning (/) edge (for example, chords 1 and 3 in Figure 1.5), and triads related by an L transformation share a left-leaning (\) vertical edge (for example, chords 1 and 4 in Figure 1.5).[90] These are not the only analytical graphs defined by neo-Riemannian theorists; others include the 'Chicken-Wire Torus', 'Cube Dance', 'Power Towers' and '4-Cube Trio'. These graphs map non-tonal music (instead of the *Tonnetz* triads) differently, using Tristan genus chords, trichords, and diminished seventh chords.[91]

Neo-Riemannian theory has already been applied to analyses of music similarity, specifically in using the *Tonnetz* and parsimonious voice-leading. Much of this work has grown out of computational research which aims to detect harmonic cycles. Distance is the critical metric for measuring similarity on the *Tonnetz*, by looking at whether two chord progressions apply equally minimal voice-leading. Andrew Milne and Simon Holland

89. Gollin, "Some Aspects of Three-Dimensional 'Tonnetz'."; Cohn, "Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective."

90. Cohn, "Neo-Riemannian Operations, Parsimonious Trichords and their "Tonnetz" representations."

91. Jack Douthett and Peter Steinbach, "Parsimonious Graphs: A Study in Parsimony, Contextual Transformations, and Modes of Limited Transpositions.," *Journal of Music Theory* 42, no. 2 (1998): 241–263; Cohn, *Audacious Euphony: Chromaticism and the Triad's Second Nature.*

(2016) define similarity as two chords that have a small distance between them.[92] One counts the number of transformations (P, R, L and D) that it takes to traverse from one chord to another on the *Tonnetz* — for example, for C major to F major the 'transformational distance' is two (see Figure 1.5, triads 1–5), as we move from C major to A minor (using the R transformation) and then to F major (using the L transformation).[93] They emphasise that for this approach to work it is crucial to count the minimal number of transformations it takes to move between two chords, rather than any cycle (for example C major to F major going via C minor, A♭ major and then F major).

Similarly, Dmitri Tymoczko (2009) also measures triadic distance using the *Tonnetz* — assuming a region (triads) that share an edge are a single unit apart.[94] This means that the three core transformations (P, R, and L) are a single unit from each other, the difference being that Tymoczko would view chords related by the D transformation as two units apart.

Milne and Holland built upon their *Tonnetz* approach to incorporate voice leading and minimal voice leading vectors.[95] They translate each pitch that makes up a triad into a mathematical pitch vector — for example, C4 is given the value 60 with each semitone up or down moving up or down 1, so C♯4 would be 61, D4 62 etc.[96] To calculate the voice-leading vector (v) you subtract the pitch vectors of the first triad (x) from the second triad (y) (i.e., $v = y - x$).[97] If we return to our example of C major to E major, this would equal:

$$
\begin{aligned}
&Cmajor = (C4, E4, G4)\\
&Cmajor = x\\
&x = (60, 64, 67)\\
&Emajor = (B3, E4, G\sharp 4)\\
&Emajor = y\\
&y = (59, 64, 68)\\
&v = y - x\\
&v = (-1, 0, 1)
\end{aligned}
$$

92. Andrew J. Milne and Simon Holland, "Empirically Testing Tonnetz, Voice-Leading and Spectral Models of Perceived Triadic Distance.," *Journal of Mathematics and Music: Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance* 10, no. 1 (2016): 59–85.

93. Ibid.

94. Dmitri Tymoczko, "A Computer Aid for Schenkerian Analysis," In *Proceedings of the 2nd international conference of Mathematics and Computation in Music* (New Haven), 2009, 258–273.

95. Dmitri Tymoczko, *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice* (New York: Oxford University Press, 2011).

96. In this thesis, I use scientific pitch notation. The first letter is the note, the number is the octave, with 4 being the octave of middle C (middle C is C4). So here A3 is the A below middle C and A4 the A above middle C.

97. Milne and Holland, "Empirically Testing Tonnetz, Voice-Leading and Spectral Models of Perceived Triadic Distance."

This approach takes into account not just the harmonic similarity between the chords, but also the inversion and octave of the chords — the vector would be significantly higher if our E major chord was not inverted (B3, E4, G♯4) as the values compare each vector (or pitch) with the equivalent vector in the other chord.

Carol Krumhansl has extensively researched the use of neo-Riemannian's transformations in music similarity, and whether chords related by a single transformation are audibly similar.[98] Similarly to distance measuring approaches already discussed, Krumhansl asserts that pitch proximity and single tone shifts aid in the independent psychological reality of neo-Riemannian transformations.[99] The transformations D, P, L, and R were particularly crucial in non-musicians' perceptions of a key (region) and the judgement of chord tension. Participants were less likely to judge tension when chords are related by a single transformation because, Krumhansl asserts, this makes them perceptually more similar.[100]

1.3.3 *Formenlehre*

Formenlehre, the traditional branch of music theory that concerns itself with the study of musical form, expresses a connection between pieces of music that have a similar structure.[101] To segment a piece of music, one splits the piece of music into elements of repetition, contrast, and variation.[102] These distinct-continuous time-spans are compared to one another by thematic or harmonic content.[103] We observe the architectural form of a piece at the most macro-level (for examples of some of the most popular western forms see Table 1.3).[104]

A letter-structure analysis can be employed in order to determine the architectural form of a piece, for example, sonata form can be expressed as ABA′ by using simplified macro-letter label analysis. Each letter represents and distinguishes a structural segment of the

98. Carol L. Krumhansl, "Tonal Hierarchies and Rare Intervals in Music Cognition," *Music Perception: An Interdisciplinary Journal* 7, no. 3 (1990): 309–324; Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations."

99. Krumhansl, "Tonal Hierarchies and Rare Intervals in Music Cognition"; Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations."

100. Krumhansl, "Tonal Hierarchies and Rare Intervals in Music Cognition"; Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations."

101. James Hepokoski, "Sonata Theory and Dialogic Form," in *Music Form, Forms, Formenlehre: Three Methodological Reflections*, 2nd, ed. Pieter Bergé (Leuven: Leuven University Press, 2010), 41–46; Nicholas Cook, "Musical Form and the Listener," *The Journal of Aesthetic and Art Criticism* 44, no. 1 (1987): 23–29.

102. William E. Caplin, *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart and Beethoven* (New York: Oxford University Press, 1998).

103. Emilios Camboropoulos, "Musical Parallelism and Melodic Segmentation: A Computational Approach," *Music Perception: An Interdisciplinary Journal* 23, no. 3 (2006): 249–268.

104. A form built upon thematic material which moves between the different parts of the music. To use alphabetical formats to explain this form, one could label the different thematic material, subject (A), answer (B), counter-subject (A‘), free part (C — then a different letter for each time this free part is different). Overall this would leave an individual structure per line for example the bottom part might be; AA‘CDA, middle part: BA‘EF, and top part AGA‘

Form	**Structure**
Fugue	Contrapuntal Form
Theme and Variation	AA'A"A"'etc.
Binary	AB
Ternary	ABA
Rondo	ABACADA etc.
Sonata Form	Exposition, Development, Recapitulation (= approximately AB – X – AB)

Table 1.3: The common musical forms in western music. Each form has its structure denoted in rough alphabetical labelling, representing repetitions, contrasts and variations — see footnote n.101 for a discussion of the difficulties in completing this kind of analysis on a piece in contrapuntal form.

piece, with the letters A and B denoting distinct (un-similar) segments.[105] To represent sections of similarity, but not 'sameness', a prime symbol is used to show a repetition with contrast (A'). As noted previously, these segments are categorised using either thematic or harmonic content; in this case, ABA' represents a thematic analysis.[106] Using William Caplin's formal functions, we can take this further and identify a 'function' for each section, in our sonata form example, making A the exposition, B the development, and A' the recapitulation.[107]

Form can also be determined at a meso- or micro-level, observing the distinct themes within a segment (meso-level) or, even further, the components that make up a particular theme (micro-level). Returning to our sonata form example, at a meso-level we could observe the exposition's (A) thematic make up of the primary theme a, and a subordinate theme b.[108] At a micro-level, we can analyse the composition of the subordinate theme (b) as ab, made up of an antecedent (a) and consequent (b).[109]

A fundamental weakness of using letters to categorise form is the subjective interpretations of repetition, contrast, or variation.[110] The use of a prime symbol to denote variation is utilised to varying degrees between analysts, with individuals weighing the variation versus contrast between sections differently. David Huron (2013) emphasises this in his research into human perception of musical form,[111] which focuses on the importance of 'habituation' — the brain's process for identifying what we have already experienced. Through this research, Huron found that small levels of similarity could cause the same decreased responsiveness seen in habituation. He identified that the amount of variation in a piece of music that an individual's brain perceives as 'still the same' varies, and in turn, affects a person's perception of form.

105. Caplin, "What are Formal Functions?"
106. Smith, "Musical Form and Fundamental Structure: An Investigation of Schenker's 'Formenlehre'."
107. Ibid.
108. Hepokoski, "Sonata Theory and Dialogic Form."
109. Ibid.
110. Caplin, "What are Formal Functions?"
111. David Huron, "A Psychological Approach to Musical Form: The Habituation-Fluency Theory of Repetition.," *Current Musicology* 96, no. 1 (2013): 7–35.

Harmonic:	A	B	
Sonata Form:	**Exposition**	**Development**	**Recapitulation**
Thematic:	A	B	A

FIGURE 1.6: The ways in which you can represent a sonata movement, including harmony (top row) and thematic material (bottom row).

Significant debate also surrounds the balance in the related importance of harmonic and thematic musical features in determining form. If an analyst were to utilise harmonic content alone to denote their labelling system, they would observe repetitions, contrasts, and variations in changes of keys or chord progressions.[112] In contrast, using thematic content, sections of a piece are categorised by associating their relationship to a recognisable melodic line, or in turn the introduction of a new theme (or thematic group).[113] Analysts disagree as to whether harmonic or thematic material should take precedence; famously, Heinrich Schenker felt that harmonic content was most important, disagreeing with William Caplin, who argued for the prominence of thematic ideas.[114] Once again returning to our sonata form example, the thematic approach of detailing ABA′ (or ABA-dependent on the similarity of the return of the first thematic group in the recapitulation), is different from the formal structure determined by using harmonic content. Sonata form, when observed from a tonal viewpoint, can be seen in two parts AB — with the tonic moving to the dominant, and then the dominant to the tonic (see Figure 1.6). Therefore, the fundamental formal analysis of the piece changes based on the approach chosen.

Adam Ockleford (2004) investigated our processing of internal similarity, salience, deviation, categorisation, and schematisation in determining the musical form of a piece.[115] Building on his own 'zygonic' theory, Ockelford emphasised that pieces of music unfold over time, i.e. musical material is derived from other aspects of the piece. For example, an entire work can grow from a four-note motive, such as that which starts Beethoven's Fifth Symphony. Similarly, Schenker's favourite example of organicism details how an ascending third progression is a seed for Bach's 12 short preludes.[116] In turn, Ockelford

112. Dahlhaus et al., "Harmony."
113. Michael Kennedy, "Thematic Material," in *The Oxford Dictionary of Music*, 2nd (Oxford University Press, 2006).
114. Smith, "Musical Form and Fundamental Structure: An Investigation of Schenker's 'Formenlehre'"; Caplin, "What are Formal Functions?"
115. Adam Ockelford, "Implication and Expectation in Music: A Zygonic Model," *Psychology of Music* 31, no. 1 (2006): 81–142.
116. Joseph Lubben, "Schenker the Progressive: Analytic Practice in 'Der Tonwille'," *Music Theory Spectrum* 15, no. 1 (1993): 59–75.

Ternary:	**A**	**B**	**A**
Binary:	**A**	**B**	
Rounded Binary:	**A**	**Ba**	

FIGURE 1.7: The differences between ternary, binary and rounded binary form.

argues that this perception of internal similarity and derivation, is central to our perception of musical form.

Another weakness of the categorisation of form arises from the lack of clarity between the definitions of certain musical forms; for example, the supposed difference between 'ternary' (ABA) and 'rounded binary' (ABa) (see Figure 1.7). 'Ternary' form refers to the musical structure ABA, where the thematic/harmonic material of A returns again after a differing thematic/harmonic structure in the middle (B). In contrast, 'rounded binary' specifically uses binary form (AB), but within the B section there is a return of the thematic material of the A section to conclude it. The lowercase 'a' denotes a shorter return of the musical material than the uppercase 'A' of ternary form. Nevertheless, we might question at what point does a rounded binary form become ternary form how much shorter does this return have to be, or how different from the initial material, to make it ABa instead of ABA?

1.3.4 Schenkerian analysis

Heinrich Schenker's late work *Der freie Satz* (Free Composition, 1935), maintained the idea that each diatonic piece of music is the 'composing out' of the 'chord of nature' (i.e. the major chord formed by the first five partials of any given fundamental), and that this is the fundamental unifying element of music.[117] In practice, the 'chord of nature' is the tonic chord of a piece of music, which emphasises the apparent simplicity, and naturalness of tonal music — with all musical works being elaborations of the tonic chord.[118]

117. Matthew Brown, *Explaining Tonality: Schenkerian Theory and Beyond* (New York: University of Rochester Press, 2005).
118. Cook, *A Guide to Musical Analysis*; Nicholas Cook, "Epistemologies of Music Theory," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2002), 78–105; Phillip B. Kirlin, "A Probabilistic Model of Hierarchical Music Analysis" (PhD diss., 2014).

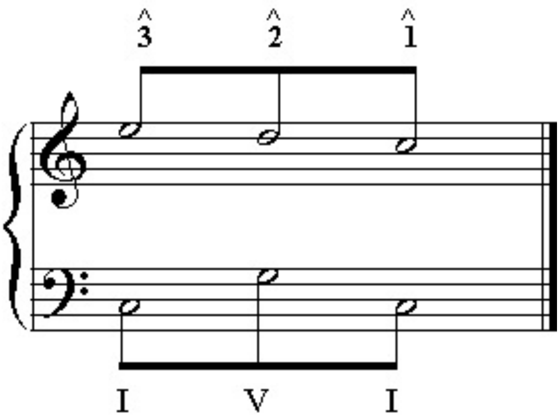


FIGURE 1.8: The descent from $\hat{3}$ Ursatz defined by Schenker in *Der freie Satz*

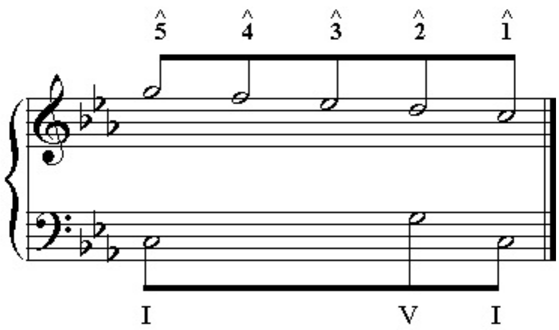


FIGURE 1.9: The descent from $\hat{5}$ Ursatz defined by Schenker in *Der freie Satz*.

Schenker introduced the idea that all musical compositions have one of three 'deep structures' named the 'Ursatz'. These three Ursatz-configurations feature an 'Urlinie', or top line, which has one of three descents from an initial tone (the 'Kopfton'): most commonly the descents from $\hat{5}$ (Figure 1.9) or $\hat{3}$ (Figure 1.8), to the final tone which is always the tonic $\hat{1}$. The other, less frequent descent, is the octave descent from $\hat{8}$. The selection of the piece's Kopfton can make a considerable difference to the analysis of the piece, and sometimes the Kopfton is not clear and can be hidden by an 'initial ascent', or disguised by a 'cover tone'.[119]

To grow through the hierarchical levels of Schenker's structural analysis, a composer performs diminutions on notes to reduce the time-span they govern.[120] Therefore, by

119. Allen Cadwallader and David Gagne, *Analysis of Tonal Music: A Schenkerian Approach* (New York: Oxford University Press, 2011).
120. Forte and Gilbert, *Introduction to Schenkerian Analysis*.

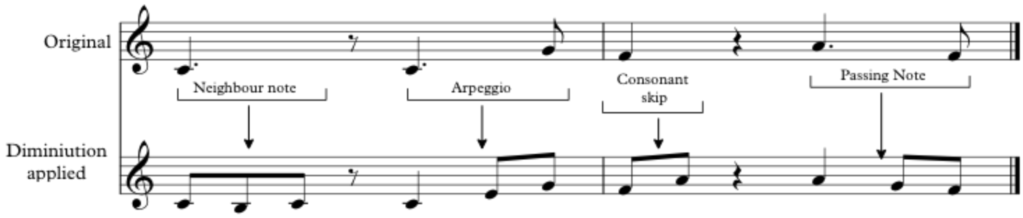


FIGURE 1.10: A variety of the different diminutions detailed by Schenker. Including Neighbour note, Arpeggio, Consonant Skip and Passing note. The top stave shows some music governing the same span of time, without the use of diminutions, the second stage shows the result of applying the different diminutions.

performing different diminutions, we reach a different foreground from the same Ursatz. There are a variety of different diminution techniques that can be used, including the passing note, neighbour note, consonant skip, and arpeggiation, to name a few (Figure 1.10). For this process of composing out the Ursatz, Schenker also uses the term 'prolongation'.[121] Prolongation is where a note governs a span of music without necessarily constantly sounding.[122] It explains how a note governs the span of music, even when it has been a part of the diminishing process (Figure 1.10). For example, in the neighbour-note example of Figure 1.10, the note C4 still governs the space of the dotted crotchet, even when it has been transformed through the diminishing process to be C4, B3, C4. Interestingly, Steve Larson (1997) discusses the human ability to hear prolongations, asserting that listeners will hear different levels of prolongations, possibly influenced by their musical training.

As Schenker sees musical works as connected in such a fundamental way, his theory is successful in showing the similarity between pieces through concepts such as the Ursätze, prolongations, linear progressions, and the hierarchical structure.[123] The distance that a piece moves away from the Ursatz, provides a comparable method for similarity between tonal pieces of music; similarity is present when the foreground of a piece is closer to the background structure. Stephen McAdams and Daniel Matzkin (2001) explored at what level of the piece's structural hierarchy the differences in the music reduced their audible similarity. They found that pieces that retained the same underlying structure, with most differences existing on the surface level, were perceived as significantly more similar

121. William Drabkin, "Prolongation," in *Grove Music Online*, ed. Stanley Sadie (Oxford University Press, 2001), http://www.oxfordmusiconline.com/subscriber/article/grove/music/22408.

122. Forte and Gilbert, *Introduction to Schenkerian Analysis*; Edward R. Pearsall, "Harmonic Progressions and Prolongation in Post-Tonal Music.," *Music Analysis* 10, no. 3 (1991): 345–355; Drabkin, "Prolongation."

123. Steve Larson, "The Problem of Prolongation in 'Tonal' Music: Terminology, Perception, and Expressive Meaning.," *Journal of Music Theory* 41, no. 1 (1997): 101–136.

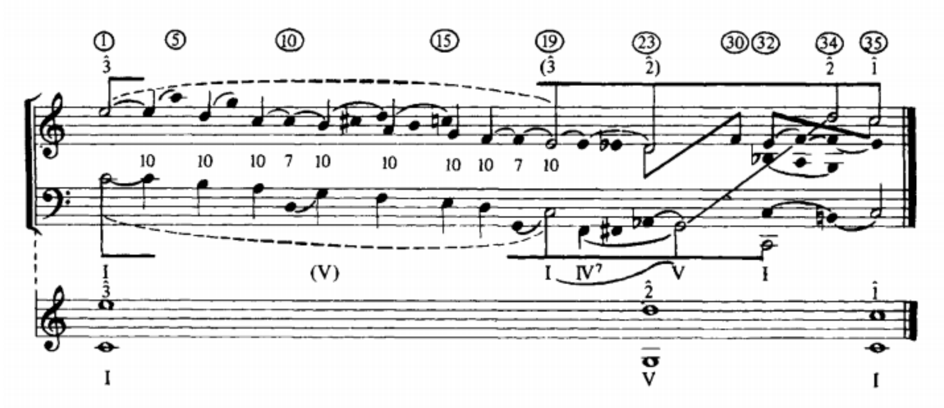


FIGURE 1.11: An example of the 7–10 linear progression from Allen Forte and Steven E. Gilbert's analysis of Bach's Prelude in C from Allen Forte and Steven E. Gilbert, *Introduction to Schenkerian Analysis*: p. 202, ex. 176.

than those where the differences existed in the background structure.[124] Leonard Meyer (1956), on the other hand, suggests that the prolongation of an Ursatz is not perceived, no matter your musical training, although he offers no evidence for this assertion.[125] In his opinion, the further we move away from the Ursatz the less likely we are to perceive its prolongation, meaning it does not affect audible similarity. He concludes by denying the perceptual significance of the Ursatz.

Allen Cadwallader and William Pastille (1992) have also discussed similarity in relation to Schenker's hierarchical structure, and specifically linear intervallic patterns.[126] Linear intervallic patterns are a voice-leading technique made up of 'successive recurrent pairs of intervals formed between the descant and the bass (outer voices)'.[127] Common patterns include the 7–10 and 10–10 intervals. Figure 1.11 provides an example of the 7–10 linear intervallic pattern in bar 10 of Bach's Prelude in C, where the seventh suspension resolves onto a tenth in the bass. Cadwallader and Pastile discuss the importance of where in the hierarchical structure a linear intervallic pattern is found, placing greater emphasis on the motives that have emanated from close to the Ursatz. Similarly, to McAdams and Matzkin (2001) and Larson (1997), Cadwallader and Pastile suggest that the elements of similarity found close to the Ursatz are important components for determining similarity between pieces.

124. Stephen McAdams and Daniel Matzkin, "Similarity, Invariance, and Musical Variation.," *Annals of the New York Academy of Sciences* 930, no. 1 (2001): 62–76.
125. Leonard Meyer, *Emotion and Meaning in Music* (Chicago: The University of Chicago Press, 1956).
126. Allen Cadwallader and William Pastille, "Schenker's High-level Motives," *Journal of Music Theory* 36, no. 1 (1992): 119–148.
127. Forte and Gilbert, *Introduction to Schenkerian Analysis*, p.83.

1.4 Computational Approaches for Music Theory

The advancements in the encoding of sheet music, discussed earlier in this chapter, have made computational score-based analysis possible. However, little work has been done to investigate using this computational format and its analysis in MIR. This section will provide a literature review of the current endeavours of computational music analysis for the theories discussed in the previous section, namely Riemannian, neo-Riemannian, *Formenhlre*, and Schenkerian analysis. If applied computationally, these analytical methods could provide a solution to the current limitations of audio analysis in extracting high-level musical features.

1.4.1 Computational approaches for Riemannian Theory

Though no significant approaches for completing a computational Riemannian analysis exist, research has explored extracting harmonic labelling systems from audio files and symbolic representation. The most significant obstacle in completing such analyses is in how to determine which notes are harmony notes, and which are non-harmony notes. Therefore, most computational harmonic analysis has focused on low-level harmonic analysis (e.g. how to extract the harmony notes from an audio file), including using a chromagram,[128] and the constant-Q spectrum.[129] Mapping these low-level features to different modelling techniques enables the extraction of the harmony notes, such as

128. G. Wakefield, "Mathematical Representation of Joint Time-Chroma Distributions," In *Proceedings of the International Symposium on Optical Science, Engineering and Instrumentation* (Denver) 99 (1999): 18–23; Rodger Shepard, "Circularity in Judgements of Relative Pitch," *Journal of the Acoustical Society of America* 36, no. 1 (1964): 23–46; Juan Pablo Bello, Giuliano Monti, and Marnk B. Sandler, "Techniques for Automatic Music Transcription," In *Proceedings of the 11th International Society of Music Information Retrieval Conference* (Massachusetts), 2000, 23–25.

129. S. Hamid Nawab, Salma Ayyash, and Robert Wotiz, "Identification of Musical Chords using Constant–Q Spectra," In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP* (Massachusetts), 2001, 3373–3376; Takuya Yoshioka et al., "Automatic Chord Transcription with Concurrent Recognition of Chord Symbols and Boundaries," In *Proceedings of the 5th International Society for Music Information Retrieval (ISMIR) Conference* (Barcelona), 2004, 3373–3376; Juan Pablo Bello and Jeremy Pickens, "A Robust Mid-Level Representation for Harmonic Content in Music Signals," In *Proceedings of the 6th International Society for Music Information Retrieval (ISMIR) Conference* (London), 2005, 304–311; Matthias Mauch, Katy Noland, and Simon Dixon, "Using Musical Structure to Enhance Automatic Chord Transcription," In *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference* (Utrecht), 2009, 231–236; Yizhao Ni et al., "An end-to-end Machine Learning System for Harmonic Analysis of Music," In *Proceedings of the IEEE Transactions on Audio, Speech and Language Processing* 20, no. 6 (2012): 1771–1783; Ruofeng Chen et al., "Chord Recognition using Duration-Explicit Hidden Markov Models," In *Proceedings of the 13th International Society for Music Information Retrieval (ISMIR) Conference* (Portugal), 2012, 445–450.

template matching,[130] and hidden Markov models.[131] This field of MIR is called ACE or Automatic Chord Extraction.[132] A recent development in this field has seen researchers use human annotator disagreement to improve ground-truth datasets, enabling ACE algorithms to act in a more human-like manner.[133]

High-level harmonic analysis (such as Riemannian theory and functional labelling of harmony) relies on a high level of musical intuition and knowledge. A sufficient coding of this intuition and knowledge is yet to exist, meaning the development of accurate MIR systems is limited. The current computational work, forefronted by the work of José Pedro Magalhães and W. Bas de Haas, relies on the input of symbolic chord labels. Their project, 'Harmonic Analysis and Retrieval of Music with Type-level Representations of Abstract Chord Entities (HARMTRACE)',[134] can automatically derive the harmonic function (e.g. tonic, dominant) of a chord in relation to its home key (provided by the user) from a set of chord labels (e.g. C major, C minor).[135] Although at an early stage, this work has demonstrated the possibility of developing MIR tools that can determine the harmonic function (from extracted chord labels). When combined with low-level harmonic extraction tools, such as the developments of ACE algorithms discussed above, a computational Riemannian analysis might be possible.

The recent work of Tsung-Ping Chen and Lin Su (2018) presents a more thorough system for harmonic function recognition.[136] Their approach utilises the BPS-FG dataset — a dataset which includes annotations of the first movement of Beethoven's Piano Sonatas with annotated functional harmony. Similarly to HARMTRACE, Chen and Su are referring to functional harmony in terms of providing chord labels (e.g. `I`, `ii`...)

130. Laurent Oudre, Yves Grenier, and Cédric Févotte, "Chord recognition using measures of fit, chord templates and filtering methods," In *Proceedings of the IEEE Workshop on Applied Signal Processing and Audio Acoustics*, 2012, 9–12; W. Bass . de Haas, José Pedro Magalhães, and Frans Wiering, "Improving audio chord transcription by exploiting harmonic and metric knowledge," In *Proceedings of the 13th International Society of Music Information Retrieval (ISMIR) Conference* (Portugal), 2012, 295–300; Giordano Cabral et al., "Automatic X Traditional Descriptor Extraction: The Case of Chord Recognition," In *Proceedings of the 6th International Society of Music Information Retrieval (ISMIR) Conference* (London), 2005, 444–449; S. Kullback and R. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics* 22, no. 1 (1951): 79–86.

131. A hidden Markov model (HMM) is a statistical model that observes a randomly changing system. The system modelled is assumed to be a state of events, where the state of the previous event determines the probability of each event. An HMM specifically has unobservable or hidden states.Takuya Fujishima, "Realtime Chord Recognition of Musical Sound: A System using Common Lisp Music," In *Proceedings of the International Computer Music Conference* (Beijing), 1999, 464–467

132. I refer the reader to Matt McVicar et al., (2014) for further discussion on the development of ACE.

133. Matt McVicar et al., "Automatic Chord Estimation from Audio: A Review of the State of the Art," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* 22, no. 2 (2014): 556–575; Hendrik Vincent Koops, "Computational Modelling of Variance in Musical Harmony" (PhD diss., Utrecht University, 2019); Selway et al., "Explaining Harmonic Inter-Annotator Disagreement using Hugo Riemann's theory of 'Harmonic Function'"; Hendrik Vincent Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music," *Journal of New Music Research* 48, no. 3 (2019): 232–252.

134. Joseé Magalhães and W. Bas de Haas, "Functional Modelling of Musical Harmony: an Experience Report," *The ACM Special Interest Group on Programming Languages Notices* 46, no. 9 (2011): 156–162.

135. Ibid.

136. Tsung-Ping Chen and Li Su, "Functional Harmony Recognition of Symbolic Music Data with Multi-Task Recurrent Neural Networks," In *Proceedings of the 19th International Society for Music Information Retrieval (ISMIR) Conference* (Paris), 2018, 90–97.

for chords within a key (the chords' relation to the tonic, subdominant and dominant), rather than in the Riemannian sense. For this, Chen and Su use recurrent neural networks, with bidirectional long short-term memory units to model functional harmony.[137] Though this work does not use a full score, it uses piano-roll segments (a visual representation of MIDI data, which represents pitch, length, and velocity of notes) as its input, and extracts the key and chord-labels computational before positing a Roman numeral analysis.

Christopher W.M. White and Ian Quinn (2018) use a data-driven approach; instead of applying functional theories, they create models of harmonic function from the corpora.[138] They presented their work as a means to explore whether the standard harmonic theory's three-function model (the tonic, dominant and subdominant) would be an accurate model for different genres of composition. White and Quinn used three corpora: the Kostka-Payne,[139] the McGill Billboard,[140] and Bach chorales. The Bach chorale corpus was the only one to produce the three-state functional model of music, though the authors argued in preference of a 13-state model due to the third state's 'messy subdominant'. However, they did find that for each corpus two of the three standard functions were significantly prominent: tonic and dominant for the Kostka-Payne and Bach chorale corpora, and the tonic and subdominant for the McGill Billboard corpus. Interestingly, they observed that the respective two functions for each corpus acted as essential pillars of the functional system.[141] The change between these pillars/functions were prolonged by adding in functions such as the pre-subdominant/or pre-dominant.

1.4.2 Computational approaches for neo-Riemannian theory

Research by Jonathan Bragg, Elaine Chew and Stewart Shieber (2011) has begun to bridge the gap between labelling algorithms and high-level harmonic computational analysis. However, their system still requires the input of a string of triads, which

137. A neural network is a group of algorithms that aims to find relationships in the data through a process that mimics how the human brain works. One class of neural networks is recurrent neural networks; this features connections between nodes in a directed graph, that exhibits temporal dynamic behaviour (bi-directional, being of both directions). Long short-term memory is one recurrent neural network architecture, this is bi-directional (having both feedback and feed-forward connections), this can process entire sequences of data, along with single data points.

138. Christopher Wm White and Ian Quinn, "Chord Context and Harmonic Function in Tonal Music," *Music Theory Spectrum* 40, no. 2 (2018): 314–335.

139. Kostka-Payne corpus contains musical examples from Kostka and Payne's textbook on tonal harmony. Stefan Kostka and Dorothy Payne. *Tonal Harmony.* 2008.

140. John Ashley Burgoyne, Jonathan Wild and Ichiro Fujinaga introduced the *Billboard* dataset in 2011, which contains chord labels for songs from the Billboard 'Hot 100' music charts, the definitive weekly ranking of the most popular songs in North America Eric T. Bradlow and Peter S. Fader, "A Bayesian Lifetime Model for the 'Hot 100' Billboard songs," *Journal of the American Statistical Association* 96, no. 454 (2001): 368–381; John Ashley Burgoyne, J. Wild, and I. Fujinaga, "An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis.," in In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami, 2011), 633–638.

141. White and Quinn, "Chord Context and Harmonic Function in Tonal Music."

represent the harmonic analysis of a piece.[142] To detect the harmonic cycle and substrings described by neo-Riemannian theorists, such as LP, RP, LRP, or LR, they apply a noisy channel model to this string of triads,[143] implemented with a weighted finite-state transducer.[144] In their article, Bragg et al. identify three reasons why automating the detection of neo-Riemannian cycles is useful: (1) to formalise a rigorous definition of cycles, (2) to facilitate a more comprehensive study of those cycles, and (3) to enable a critique of neo-Riemannian theory.

Similarly to Riemannian theory, very little research to date has explored creating computational approaches for neo-Riemannian theory. Considering neo-Riemannian theory's substantial mathematical background, one would anticipate an aptness of the theory for computation. The work of researchers such as Dmitri Tymoczko and David Lewin exemplifies this mathematical foundation.[145] In turn, this has led to some readily available computer-aided neo-Riemannian analysis tools, such as HexaChord,[146] and Open Music.[147] The first, HexaChord, takes a MIDI file and outputs a three-dimensional *Tonnetz*.[148] A variety of musical representation spaces are selectable, including chromatic circles, circles of fifths, and voice leading proximity. With the aim of aiding music analysis, the visualisation of a piece on the *Tonnetz* moves in real-time alongside a playback of the MIDI file.[149] The second, OpenMusic, is a visual/graphical programming language using LISP. It enables the music theorist to visualise musical properties of a piece in a geometric way, using algebraic structures. Similarly to HexaChord, OpenMusic enables the extraction of chord progressions and visualisation of these progressions on a *Tonnetz*.[150]

142. Jonathan Bragg, Elaine Chew, and Stuart Shieber, "Neo-Riemannian Cycle Detection with Weighted Finite-State Transducers.," In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 399–404.

143. A noisy channel model assumes there is some error in our input and cleans the data back to its assumed 'error-free' version, using a prediction of plausible alternatives — often this model is used in spell checking.

144. A finite-state model stores the state or a record of something at a specific time. It can operate on an input to change the status and/or cause an action or output to take place. The defining factor which makes a state machine finite is where it has a limited, or 'finite', number of possible states. This method is used often in language and speech processing. Bragg et al. use a weighted finite-state model, which enables the annotation of the likelihood that the triad is a part of a cycle.

145. Tymoczko, *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*; David Lewin, *Generalized Musical Intervals and Transformations* (Oxford: Oxford University Press, 2010).

146. https://www.louisbigo.com/hexachord

147. https://www.repmus.ircan,fr/openmusic/home

148. Louis Bigo et al., "Computation and Visualisation of Musical Structures in Chord-Based Simplicial Complexes," In *Proceedings of the International Conference on Mathematics and Computation in Music* (Berlin), 2013, 38–51.

149. Ibid.

150. Jean Bresson, Agon Carols, and Gérard Assayag, "OpenMusic: Visual Programming Environment for Music Composition, Analysis and Research," In *Proceedings of the ACM MultiMedia (OpenSource Software Competition),* (Arizona), 2011,

1.4.3 Computational approaches for *Formenhlre*

Most research into computational approaches to form have used the music of Johann Sebastian Bach, other baroque pieces, or songs from popular music. This research has mostly focused on form from the perspective of thematic similarity, on a small scale for example, Mathieu Giraud, Richard Grautt and Florence Levé (2016) developed an algorithm to extract the episodes from the fugues in Bach's first book of the Well-Tempered Clavier.[151] Similarly, Gabriel Sargent et al. in their MIREX competition submission (2011), worked on computationally extracting the episode from a fugue to determine its structure.[152] In popular music, computational approaches have also focused on picking out thematic units; for example, the work of Mark Levy, Mark Sandler, and Michael Casey (2006), and Masataka Goto (2003).[153] Their research looked at finding the 'most repeated segment' (likely the chorus), using self-similarity matrixes.[154] In both musical genres, the research has looked only on a small scale at picking out the main subjects, focusing on the small-scale formal role of thematic units, thus avoiding the issue of the balance between harmonic and melodic articulations, which come into play in larger formal units.

Full-scale segmentation of form remains a challenging problem for feature extraction. One approach utilised for this, has been self-similarity matrices (e.g. Namunu Maddage (2006),[155] Lu, Wang and Zhang (2004),[156] Paulus and Klapuri (2006),[157] and Chai (2006)).[158] Though this approach shows promise in terms of the extraction of small-scale form and finding the 'most repeated segment',[159] it appears not to be as straight forward when extending this search to full-piece segmentation. For example, research by Lie Lu,

151. Jennifer Giraud, Richard Groult, and Levé Florence, "Computational Analysis of Musical Form," in *Computational Music Analysis*, ed. David Meredith (New York: Springer, 2016), 113–136.
152. Gabriel Sargent et al., "A Music Structure Inference Algorithm Based on Symbolic Data Analysis," *Submission to MIREX competition at 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011,
153. Mark Levy, Mark Sandler, and Michael Casey, "Extraction of High-Level Musical Structure from Audio Data and its Application to Thumbnail Generation," In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing* (France), 2006, Masataka Goto, "A Chorus-Section Detecting Method for Musical Audio Signals," In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Hong Kong), 2003, 437–440.
154. Levy, Sandler, and Casey, "Extraction of High-Level Musical Structure from Audio Data and its Application to Thumbnail Generation"; Goto, "A Chorus-Section Detecting Method for Musical Audio Signals."
155. Namunu Chinthaka Maddage, "Content-Based Music Structure Analysis" (PhD diss., 2006).
156. Lie Lu, Muyuan Wang, and Zhangm Hong-Jiang, "Repeating Pattern Discovery and Structure Analysis from Acoustic Music Data," In *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval* (New York), 2004, 275–282.
157. Jouni Paulus and Anssi Klapuri, "Music Structure Analysis by Finding Repeated Parts," In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia* (California), 2006, 59–68.
158. Wei Chai, "Semantic Segmentation and Summarization of Music: Methods based on Tonality and Recurrent Structure," *IEEE Signal Processing Magazine* 23, no. 2 (2006): 124–132.
159. Levy, Sandler, and Casey, "Extraction of High-Level Musical Structure from Audio Data and its Application to Thumbnail Generation."

Muyuan Wang, and Zhang Hong-Jiang (2004) had to rely on limiting assumptions about the form, such as pre-defining the length of a significant segment.[160]

Other approaches use low-level hidden Markov model state labelling (such as Michael Casey (2001)),[161] and histogram clustering (such as Jonathan Foote (2000)).[162] Mark Levy and Mark Sandler (2008) take a combined approach, utilising both a 40-state hidden Markov model and histogram clustering. Levy and Sandler's 'Segmenter' divides an audio file into structurally consistent segments, identified by alphabetic labelling (e.g. ABA). The method relies upon structural and timbral similarities to extract the architectural form, using a chromagram,[163] and Mel-frequency cepstral coefficients (MFCC).[164] Following this, the 40-state hidden Markov model is trained, with each of the states corresponding to a specific 'timbre type'. By clustering these states into 'segment types' using histogram clustering, 'similar' segments are associated with each other through having the same/or similar clustering of states. This approach exists as a vamp-plugin for the MIR analysis tool, Sonic Visualiser.[165] Though promising, this work has limited application in complex classical forms which feature altered repetition (e.g. A′). There is no facility for different levels of similarity, or a transparent approach to what makes something a different letter category. Similarly, the use of timbral features could also prove problematic; these are very ill-defined in MIR.

Arguably the most promising developments for the field of computational form analysis has come from the project SALAMI (Structural Analysis of Large Amounts of Music Information). This research sought to test the accuracy of a variety of form extraction algorithms, using a large (and growing) corpora of digitally recorded music.[166] A 'ground truth' dataset, created by human annotators from McGill University, the University of

160. Lu, Wang, and Hong-Jiang, "Repeating Pattern Discovery and Structure Analysis from Acoustic Music Data."
161. Michael Casey, "General Sound Classification and Similarity in MPEG-7," *Organised Sound* 6, no. 2 (2001): 153–164.
162. Histogram clustering, is a set of points clustered together on a histogram.Jonathan Foote, "Automatic Audio Segmentation using a Measure of Audio Novelty," In *Proceedings of the IEEE International Conference on Multimedia and Expo* (New York), 2000, 452–455
163. Or also known as a chroma-features are a set of profiles of different pitches (often 12)
164. MFCC's are a set of features (usually of 10–20), which describe the curve of sound in the frequency-amplitude plane. In music this is often used to describe timbre.
165. Levy, Sandler, and Casey, "Extraction of High-Level Musical Structure from Audio Data and its Application to Thumbnail Generation."
166. David De Roure, J. Stephen Downie, and Ichiro Fujinaga, "SALAMI: Structural Analysis of Large Amounts of Music Information," In *Proceedings of the 10th UK e-Science All Hands Meeting* (York), 2011, https://eprints.soton.ac.uk/271171/.

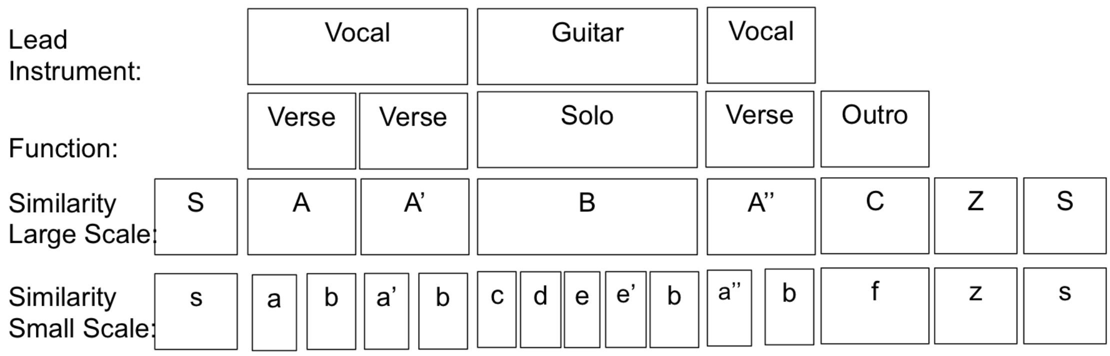


FIGURE 1.12: The different levels of the hierarchical analysis used in SALAMI. Diagram taken from Jordan B. Smith et al., 'Design and Creation of a Large-Scale Database of Structural Annotations,' In *Proceedings of the 12th International Society of Music Information Retrieval Conference* (Miami), 2011, 555–560. The piece of music is generic, not specific

Southampton, and Oxford University, enabled the evaluation of these extraction algorithms.[167] To ensure consistency in the human analysis, the researchers detailed a hierarchical annotation format based on the work of Peeters and Deruty (see Figure 1.12).[168] This hierarchical approach distinguishes not only the instrumentation of sections, but also the function, and the musical similarity of the sections with segments.

The analytical format in Figure 1.12 takes a two-layered approach, with each track identifying similar musical ideas using the traditional letter-based format discussed in Section 1.3.3. The large-scale similarity track uses uppercase letters, and the small-scale one uses lowercase. This segmentation might, for example, represent on a large scale the architectural structure of the piece, and on a small-scale the specific thematic material of each segment. The function track (which generally aligns with the large-scale segment boundaries) provides, when appropriate, a function for the section. For example, in the analysis of a classical piece, this might highlight the 'exposition', 'recapitulation' and 'development' in sonata form. For this track, the researchers used 20 pre-defined labels, including 'chorus', 'transition', 'exposition', and 'verse'. This hierarchical analysis enables us to see the similarity between pieces' forms, as all pieces fit into the same

167. Council on Library and Information Resources, "Structural Analysis of Large Amounts of Music Information (SALAMI)," CLIR, 2004, accessed April 28, 2018, https://www.clir.org/pubs/reports/pub151/case-studies/salami; De Roure, Downie, and Fujinaga, "SALAMI: Structural Analysis of Large Amounts of Music Information"; Jordan B. Smith et al., "Design and Creation of a Large-Scale Database of Structural Annotations," In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 555–560.
168. Geoffroy Peeters and Emmanuel Deruty, "Is Music Structure Annotation Multi-Dimensional? A Proposal for Robust Local Music Annotation," In *Proceedings of the 3rd Workshop on Learning the Semantics of Audio Signals* (Austria), 2009, 75–90.

segmentation structure. The SALAMI team further enabled for cross-genre comparison of form, through implementing and designing an ontology for music structure.[169]

Through evaluating these extraction algorithms, Smith et al. found that the highest accuracy of form extraction existed when combining multiple existing algorithms.[170] Therefore, the team developed an interactive visualiser interface to enable users to examine and playback individual segments of a piece, allowing them to decide which algorithms produced the most accurate formal analysis (see Figure 1.13). Another significant finding from this study was that the algorithms did not perform as well as the human annotators across their test corpus of 1400 musical recordings.[171] Though this may have been disappointing for the researchers, it raises interesting questions as to how the brain processes and understands musical structure.[172] This, again, highlights the limiting assumptions that those who create these algorithms (whether it be which musical features to use, or what length segments to examine) have to define to enable the extraction of musical form.

1.4.4 Computational approaches for Schenkerian Analysis

The largest body of computational approaches for music theory exist for Schenkerian analysis, with the work of Michael Kassler most often sighted as the earliest.[173] Kassler aimed to 'prove' Schenkerian theory, showing that music is reducible to one of three Ursatz.[174] Kassler recasts the outline of Schenker's theory in a formalised language, also using a generative approach,[175] focusing on the concept of the Ursatz and prolongation techniques.[176] In his work, Kassler uses an axiomatic logical system which assumes that a musical work derives from an axiom (a statement assumed to be true) from which

169. De Roure, Downie, and Fujinaga, "SALAMI: Structural Analysis of Large Amounts of Music Information."
170. Council on Library and Information Resources, "Structural Analysis of Large Amounts of Music Information (SALAMI)."
171. Council on Library and Information Resources, "Structural Analysis of Large Amounts of Music Information (SALAMI)"; Smith et al., "Design and Creation of a Large-Scale Database of Structural Annotations."
172. Council on Library and Information Resources, "Structural Analysis of Large Amounts of Music Information (SALAMI)."
173. Michael Kassler, "A Trinity of Essays: Toward a Theory that is the Twelve-Note Class System; Toward Development of a Constructive Tonality Theory Based on Writings by Heinrich Schenker; Toward a Single Programming Language for Musical Information Retrieval" (PhD diss., 1967); Michael Kassler, *Proving Musical Theorems I: The Middleground of Heinrich Schenker's Theory of Tonality* (The University of Sydney, 1975); Michael Kassler, "Explication of the Middleground of Schenker's Theory of Tonality," *Miscellanea Musicologica: Adelaide Studies in Musicology* 9 (1977): 72–81; Michael Kassler, "APL Applied in Music Theory," In *Proceedings of the International Conference on APL* (Texas), 1987, 209–214.
174. Heinrich Schenker, *Der Freie Satz* (Vienna: UE, 1935).
175. Kassler, "A Trinity of Essays: Toward a Theory that is the Twelve-Note Class System; Toward Development of a Constructive Tonality Theory Based on Writings by Heinrich Schenker; Toward a Single Programming Language for Musical Information Retrieval."
176. Ibid.

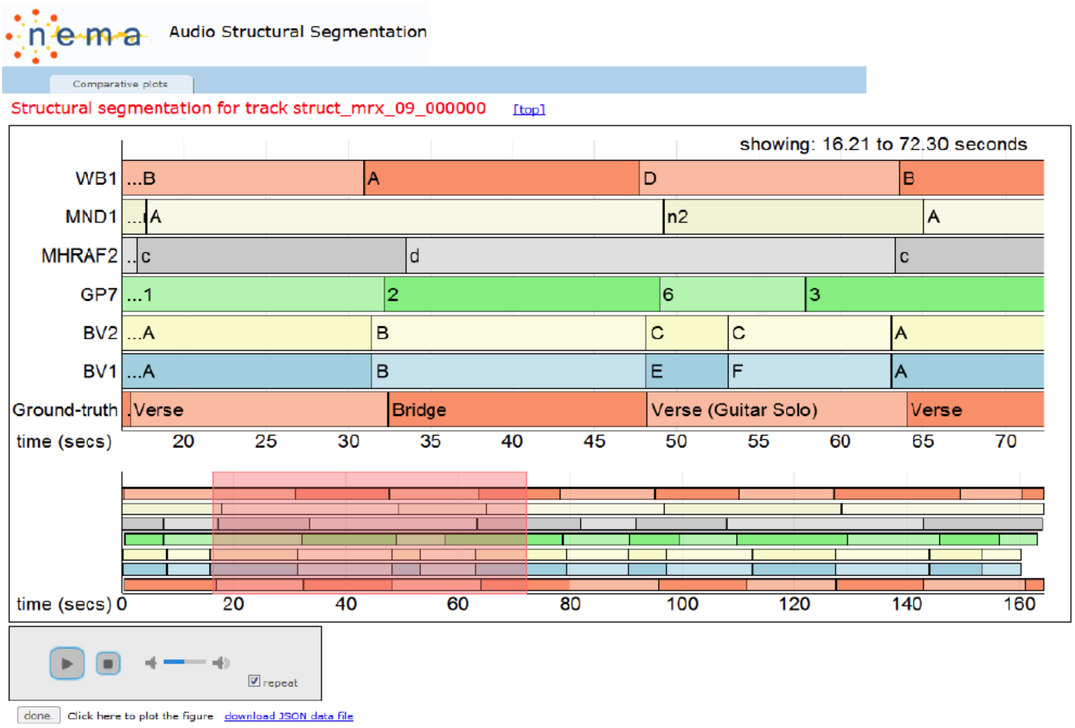


FIGURE 1.13: The visualiser created by the SALMAI project.

we can infer other statements (a theorem) through using rules of inference.[177] These 'rules of inference' are formalised versions of Schenker's prolongation techniques, as seen in Table 1.4. From this Kassler looks to assume whether a statement conforms to the defined logical system;[178] in this case, can the music be reduced to one of the Ursatz, or not?

Kassler never establishes a method for deriving a full Schenkerian analysis from the surface of a composition.[182] Instead, Kassler's work is successful in showing us how to get from a middleground reduction of a piece to a known Ursatz (i.e. we 'know' the result we are after), through performing the rules of inference in reverse.[183] Even though it fails to validate Schenker's theory, Kassler's work successfully demonstrates the possibility

177. Eric Regener, "Layered Music-Theoretic Systems," *Perspectives of New Music* 6, no. 1 (1996): 52–62; Marsden, "Generative Structural Representation of Tonal Music."; Kassler, *Proving Musical Theorems I: The Middleground of Heinrich Schenker's Theory of Tonality*.

178. Kassler used the term logistic

182. Alan Marsden, "Towards Schenkerian Analysis by Computer: A Reductional Matrix," In *Proceedings of 31st the International Computer Music Conference* (Barcelona), 2005, 247–250; Alan Marsden, "Automatic Derivation of Musical Structure: A Tool for Research on Schenkerian Analysis," In *Proceedings of the 8th International Society for Music Information Retrieval (ISMIR) Conference* (Vienna), 2007, 55–58.

183. Marsden, "Towards Schenkerian Analysis by Computer: A Reductional Matrix"; Alan Marsden, "Schenkerian Analysis by Computer: A Proof of Concept," *Journal of New Music Research* 39, no. 3 (2010): 269–289; Phillip B. Kirlin, "Using Harmonic and Melodic Analyses to Automate the Initial

Schenker's Prolongation Technique	**English Translation**	**Kassler's corresponding rule of inference**
Aufwärts-Bassbrechung[179]	Upwards Bass Arpeggiation	Rule of Bass Arpeggiation
Aufwärts-Bassbrechung	Allows for passage from C to G	Rule of Bass Ascent
Kassler's elaboration of Aufwärts-Bassbrechung	?	Rule of Bass Descent
Kassler's elaboration of Aufwärts-Bassbrechung	Transfers the Bass, of the C major Ursatz, G down an Octave	Rule of Bass G transfer
Gliederung	Structure	Rule of Articulation
Mischung	Mixture	Rule of Mixture
Nebennote	Neighbour Note	Rule of Neighbour Note Prolongation
Zug	Linear Progression (falling)	Rule of First Order Descending Progression
Zug (Ansteig)	Linear Progression (Initial Ascent)	Rule of Preliminary Ascent
Brechung	Arpeggiation	Rule of Preliminary Arpeggiation
Übergreifen	Reaching Over	Rule of Overlapping
Untergreifen	Motion from Inner Voice	Rule of Middle-Lyne Ascent ('Underlapping')
Ausfaltung	Unfolding	Rule of Unfolding
(implicit)	*[This is not an 'official' Schenkerian prolongation techniques, however Kassler has implied that this prolongation is implied by Schenker in* Der Freie Satz*]*	Rule of Transformation
Tieferlegung	Upward Octave Adjustment	Rule of Upward Octave Adjustment
Höherlegung	Downard Octave Adjustment	Rule of Downard Octave Adjustment

TABLE 1.4: Schenker's prolongation techniques and Kassler's corresponding rules of inference, highlighting which elements of Schenkerian analysis Kassler has incorporated into his computational process and how he might have adapted Schenker's techniques in his rules of inference. The table is based on the table found in Kassler[181]

for automated Schenkerian analysis. We can see Kassler's methodology of reducing or generating a piece from two notes of music, being later adopted in the work of Alan Marsden, who translates Schenker's graphs into figure trees.[184]

Stages of Schenkerian Analysis," In *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference* (Kobe), 2009, 423–428.
184. Marsden, "Schenkerian Analysis by Computer: A Proof of Concept."

Panayotis Mavromatis and Matthew Brown also took a grammar-based approach, dealing with multi-voice structures.[185] In their paper, they demonstrate the technical feasibility of expressing Schenkerien analysis in a context-free grammar. This type of grammar has a simple and effective parsing mechanism,[186] which Mavromatis and Brown suggest enables a grammar to be the basis for an automatic derivation of Schenkerian analysis from a piece of music.[187] However, this 'promise' has not been fulfilled, according to Alan Marsden, due to the 'number of rewrite rules required being preventatively large'.[188]

Stephen Smoliar and his collaborators used LISP to define a set of prolongation functions in a list structure.[189] Taking a generative approach, Smoliar focuses primarily on single note transpositions, such as neighbour notes and passing notes. In some areas, Smoliar's work reveals a more in-depth understanding of Schenker's theory than Kassler shows, specifically in his emphasis on the generation of binary vs ternary structures. However, Smoliar selects the aspects of Schenker's theory that work best for his purpose, which is a limiting factor of all computational Schenkerian analysis approaches. Concepts such as 'initial ascent' and 'reaching over', which Kassler chose to incorporate, do not feature in Smoliar's work. Still, another significant limitation of Smoliar's work is the lack of progress in automating Schenkerian analysis, requiring manual intervention from the user — by typing the function they wish to complete.[190]

Perhaps the most prominent body of research in this field comes from the work of Alan Marsden. Marsden begins his work by adapting Smoliar's model to enable a generative representation — something he believes the approach lacks.[191] Marsden's work takes the tree-like structures found in the work of Kassler and Smoliar, and represents these as directed acyclic graphs (DAGs), which tend towards binary trees.[192] Marsden's early work used dynamic programming to retrieve a single Schenkerien reduction from a matrix,

185. Panayotis Mavromatis and Matthew Brown, "Parsing Context-Free Grammars for Music: A Computational Model of Schenkerian Analysis," In *Proceedings of the 8th International Conference on Music Perception and Cognition* (Illinois), 2004, 414–415.

186. Marsden, "Schenkerian Analysis by Computer: A Proof of Concept."

187. Mavromatis and Brown, "Parsing Context-Free Grammars for Music: A Computational Model of Schenkerian Analysis."

188. Marsden, "Schenkerian Analysis by Computer: A Proof of Concept," p.271.

189. Marsden, "Schenkerian Analysis by Computer: A Proof of Concept"; Stephen W. Smoliar, Robert E. Frankel, and Stanley J. Rosenschein, "A LISP-Based System for the Study of Schenkerian Analysis," *Computer and the Humanities* 10 (1976): 21–32; Stephen W. Smoliar, Robert E. Frankel, and Stanley J. Rosenschein, "Schenker's Theory of Tonal Music — its Explication Through Computational Processes," *International Journal of Man-Machine Studies* 10, no. 2 (1978): 121–138; Stephen W. Smoliar, "A Computer Aid for Schenkerian Analysis," *Computer Music Journal* 4, no. 2 (1980): 41–59. LISP is a computer-language designed to manipulate list structures, formed of either an atom or a list. 'Atoms' in LISP, is a word for all data types that are not in an ordered pair. So for the numbers 1, 2 if the order matters they are not an atom (they are a list) but if the order does not matter they are an atom.

190. Kirlin and Utgoff, "A Framework for Automated Schenkerian Analysis."

191. Marsden, "Generative Structural Representation of Tonal Music."

192. A graph structure using nodes, with directed edges (meaning that one node can only move in one direction to the other). Being acyclic means there is no loop back to the starting node by following the directed path. Binary trees are a data structure, where a 'node' links to at most two 'child nodes'.

extracting a structural analysis for the path from the foreground to the background.[193] His matrices do not present a single reduction, dispensing with Schenker's belief of a single correct analysis. Instead, the user can select segments at each level of the reduction to produce a reduction that aligns with their analytical opinion.[194] One limitation of Marsden's approach is the requirement for a piece of music to fit into his strict matrix structure, having notes that start and finish with a new segment — often relying on pieces of music where all parts move at the same time.[195] Therefore, Marsden has yet to complete a computational analysis of an entire piece of music.

Collaborative research from Alan Marsden and Gerant Wiggins has shown how heuristics could aid Marsden's system, enabling the characterisation of some reductions as 'good' analyses.[196] Using a small corpus of six themes from Mozart piano concertos, they used a chart parser to derive a 'goodness metric' of the different analyses. This choice of repertoire is interesting, as Mozart was one of the prominent composers whose music Schenker frequently chose to analyse. Overall, Marsden and Wiggins found that combining the A* and breadth-first algorithms both extracted all possible analyses, and identified the best analysis.[197] Breadth-first enabled the search of every node to find all possible analyses, then used A* to search for the best reduction.[198]

One aspect of Marsden's approach, which appears equally valid for the whole field of computational Schenkerian analysis, is its focus on finding prolongations and reductions, ultimately leading to the Ursatz. I am not aware of any computational Schenkerian approaches that also incorporate other (admittedly secondary) features of Schenkerian analysis, such as determining linear intervallic pattern, architectural form, and parallelisms. Again, this highlights the simplification of Schenkerian analysis adopted computationally, removing the aspects of Schenker's theory that are most relevant for this thesis's exploration of music similarity. Equally, one wonders if this simplification of Schenker's approach is the reason why automated Schenkerian analysis has proved so evasive.

193. Alan Marsden, "Representing Melodic Patterns as Networks of Elaborations," *Computers and Humanities* 35 (2001): 37–54; Marsden, "Generative Structural Representation of Tonal Music."; Marsden, "Towards Schenkerian Analysis by Computer: A Reductional Matrix"; Marsden, "Automatic Derivation of Musical Structure: A Tool for Research on Schenkerian Analysis."

194. Alan Marsden, "Software for Schenkerian Analysis," In *Proceedings of 37th the International Computer Music Conference* (Huddersfield), 2011, 673–676.

195. Alan Marsden, Keiji Hirata, and Satoshi Tojo, "Towards Computable Procedures for Deriving Tree Structures in Music: Context Dependency in GTTM and Schenkerian Theory," In *Proceedings of the Sound and Music Computing (SMC) Conference* (Stockholm), 2013, 360–367.

196. Alan Marsden and Gerant A. Wiggins, "Schenkerian Reduction as Search," In *Proceedings of the Fourth Conference on Interdisciplinary Musicology* (Thessaloniki), 2008, 1–9.

197. A* is also known as best-first. The A* search algorithm attaches a judge of 'goodness' to each node. The search favours the route that has the highest probability of being the 'best'. This approach, therefore, favours finding the 'best' Schenkerian analysis, not finding all possible approaches. This search method enables us to direct our search.

198. Breadth-first search starts at the tree root, and will explore all of the neighbour nodes on the same hierarchical level, before traversing to the next level of depth.George F. Luger and William A. Stubblefield, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (Benjamin-Cummings Publishing Company, 1992); Marsden and Wiggins, "Schenkerian Reduction as Search"; Kirlin, "A Probabilistic Model of Hierarchical Music Analysis"

Parallel to this research, Philip Kirlin and his collaborators (and Marsden in a more recent publication with Keiji Hirata and Satoshi Tojo (2013)) have focused on using computational linguistics instead of tree-based structures. Marsden argued that this approach could prove more useful than his tree-based approach as context-dependency complicates the formulation of an effective computable procedure for automatically deriving trees from a piece of music; whereas, using computational linguistics allows a clarity between the structure of data and algorithms in a grammar, not possible with trees and matrices.[199] Though the work in this domain has not reached the magnitude of the work of Marsden, the developments by Kirlin and his collaborators (Paul Utgoff, David Jensen and David Thomas),[200] have led to the creation of a promising probabilistic model of Schenkerian analysis. Building on the work of Kassler, Smoliar and Marsden, Kirlin sought to overcome some of the shortcomings identified particularly in Marsden's work. These criticisms include its lack of utilising notational information through using music encoding (e.g. MusicXML), and the completion of a full Schenkerian analysis from the foreground of a piece.

From Kirlin and Jensen's 2011 paper, they adopt the use of MOPs (maximal outerplanar graphs) to represent a Schenkerian reduction defining three ParseMOP algorithms.[201] They found that their ParseMOP B had the most 'edge accuracy' (finding the correct triangulations), with most of the errors existing in the middle levels of the hierarchy (Schenker's 'middleground'), and no errors made in the background level.[202] It is not surprising that most errors were made in the middleground triangulations, because this is the most subjective part of Schenker's hierarchy. It is possible to have the same foreground and background for multiple middle-grounds. One of the main limitations of this approach is its inability to represent prolongation situations between multiple voices — therefore, the system only accepts monophonic input.[203] Also, this approach

199. Marsden, Hirata, and Tojo, "Towards Computable Procedures for Deriving Tree Structures in Music: Context Dependency in GTTM and Schenkerian Theory."

200. Kirlin, "Using Harmonic and Melodic Analyses to Automate the Initial Stages of Schenkerian Analysis"; Phillip B. Kirlin, "A Data Set for Computational Studies of Schenkerian Analysis," In *Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference* (Taipei), 2014, 213–218; Phillip B. Kirlin, "A Lesson in Analysis from Heinrich Schenker: The C Major Prelude from Bach's Well-Tempered Clavier, Book I," In *Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference* (Taipei), 2014, 213–218; Kirlin, "A Probabilistic Model of Hierarchical Music Analysis"; Phillip B. Kirlin, "Global Properties of Expert and Algorithmic Hierarchical Music Analyses," In *Proceedings of the 17th International Society for Music Information Retrieval (ISMIR) Conference* (New York), 2016, 640–646; Kirlin and Utgoff, "A Framework for Automated Schenkerian Analysis"; Phillip B. Kirlin and David D. Jensen, "Probabilistic Modelling of Hierarchical Music Analysis," In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 393–398; Phillip B. Kirlin and David D. Jensen, "Using Supervised Learning to Uncover Deep Musical Structure," In *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (Texas), 2015, 1770–1777; Phillip B. Kirlin and David L. Thomas, "Extending a Model of Monophonic Hierarchical Music Analysis to Homophony," In *Proceedings of the 16th International Society for Music Information Retrieval (ISMIR) Conference* (Málaga), 2015, 715–721.

201. Kirlin and Jensen, "Probabilistic Modelling of Hierarchical Music Analysis."

202. Kirlin and Thomas, "Extending a Model of Monophonic Hierarchical Music Analysis to Homophony."

203. Kirlin, "A Data Set for Computational Studies of Schenkerian Analysis"; Kirlin and Thomas, "Extending a Model of Monophonic Hierarchical Music Analysis to Homophony."

struggles with repeated pitches in the input music; therefore, it struggles to identify similarity, possible repeated patterns and decoration.[204]

In 2016, Kirlin defined a set of global properties by which we can discuss MOPs.[205] These include finding the height, average path length, and the left/right skewness of the MOP. These metrics enable the comparison of ground-truth data (analysts' interpretation) with the algorithmic output. These global properties could be utilised in the music similarity domain, by (for example) enabling the comparison of MOPs of multiple pieces using graph theorem-like measurements to see if 'similar' MOPs align with audible similarity.

1.5 Thesis structure and overview

This chapter opened by exploring the current approaches to music similarity within the fields of music plagiarism and music recommendation. I particularly highlighted that traditional music theory is underutilised within these applications of music similarity. An exploration of current prominent music theories (Riemannian theory, neo-Riemannian theory, *Formenhlre*, and Schenkerian analysis) provided an insight into how these theories are each useful in understanding similarity. Prominently, these theories could provide a way of extracting high-level musical features, an area in which the current MIR methods have reached their glass ceiling. As shown, there has been some computational development in using these approaches, yet little that has enabled MIR to utilise these theories within its existing approaches efficiently. This thesis, therefore, aims to understand to what end music theory can aid our understanding of audible music similarity, and to bring music theory into the debate on this topic in the MIR community. To achieve this, Study 1 (Chapter 2) examines the perceived audibility of theoretical definitions of musical similarity to focus the scope of this thesis. Study 1 shows that harmony, particularly concerning Riemannian theory, has a prominent role in listeners' perception of musical similarity. The results of this first study will narrow the scope of the thesis to focus on theories of harmony as a tool to explain music similarity.

Part II of this thesis utilises Hugo Riemann's theory of harmonic functions and aspects of prolongation from Schenkerian analysis to explain the apparent discrepancies in human-annotated harmony datasets, specifically the Chordify Annotator Subjectivity Dataset (Chapter 3), a subset of Chrodify's user data (Chapter 4), and the remaining songs from both datasets in Chapter 5. These chapters utilise Riemannian theory to observe quantitative harmonic disagreement, finding that Riemannian theory can explain a proportion of harmonic disagreement that current MIR pitch class methods have overlooked. Chapter 4 also utilises an interview study (see Chapter 4 Section 4.3) to explore how and why users of Chordify make harmonic edits. This qualitative methodology places emphasis

204. Kirlin, "A Data Set for Computational Studies of Schenkerian Analysis."
205. Kirlin, "Global Properties of Expert and Algorithmic Hierarchical Music Analyses."

on the importance of incorporating music theory into harmonic agreement measures as not only something observed in quantitative tests, but also detailed as distinguished methods for harmonic changes as described by participants.

Chapter 5 completes Part II of this thesis by discussing the 11 songs from the two datasets that have no available score. As popular music does not always have a readily available score, this chapter provides a methodology that retains Riemann's substitutions from the theory while negating the requirement for a score. This chapter finds a similar level of explainable disagreement using this non-score approach, providing an exciting potential for the ease of application of Riemannian theory in ACE music recommendation algorithms.

Building off the work of Part II, Chapter 6 presents this thesis's final study, which explores a set of predictions made from the previous chapter's findings. By asking 15 transcribers to annotate the harmony of an edited version of 'Little Bit O'Soul' by the Music Explosion (1967), this study confirms the results in Chapter 5 that a substitution-only approach is sufficient for explaining annotator disagreement, thus removing the requirement for a score. However, Chapter 6 does show that the musical score can provide additional value for explaining other causes of harmonic disagreement, such as prominent recurring aspects of the music, and disagreement occurring at points of harmonic change.

Chapter 7 concludes this thesis by providing a variety of critical findings from this, including: (1) the ability for Riemann's theory of harmonic functions to explain a higher proportion of disagreement between composers and popular musicians; (2) the *Parallele* substitution being the most common relationship between chords in harmonic disagreement; and (3) that harmonic disagreement is most likely to arise at points of harmonic change. This thesis concludes that Rieman's harmonic theory can enrich our understanding of music similarity. The potential to perform this computationally has been made possible by the work of José Pedro Magalhães and W. Bas de Hass (2011), and Tsuing-Ping Chen and Lin Su (2018), whose combined work proves the potential for computational Riemannian analysis.[206] In turn, this will improve the accuracy of MIR chord extraction algorithms, aiding our understanding of music similarity for music perception and music plagiarism.

206. Magalhães and Haas, "Functional Modelling of Musical Harmony: an Experience Report"; Chen and Su, "Functional Harmony Recognition of Symbolic Music Data with Multi-Task Recurrent Neural Networks."

Chapter 2

Audible perceptions of musical similarity in Johannes Brahms's *Variations on a Theme by Paganini*

2.1 Introduction

The notions of variation, perceptual similarity and invariance are crucial to the theme and variation formal type.[1] This form depends on the listener being able to recognise similarity, distinguish the recurrent motives, and perceive the changes that have occurred throughout a set of variations.[2] Similarities are central to the unfolding of the musical structure during the listening process,[3] enabling the listener to recognise important components of the music (those which are essential for musical memory and the experience of similarities within and between pieces of music).[4] Importantly, similarities of motive, harmony, timbre, and texture help audience members to appreciate and comprehend the music without requiring a musical education or understanding of compositional rules.[5] Due to the importance of similarity in theme and variation sets, it is not surprising that this type of musical form features prominently in music similarity perception studies.

1. McAdams and Matzkin, "Similarity, Invariance, and Musical Variation."

2. Alan Marsden, "Interrogating Melodic Similarity: A Definitive Phenomenon or the Product of Interpretation?," *Journal of New Music Research* 41, no. 4 (2012): 323–335.

3. Lawrence Zibikowski, *Conceptualizing Music: Cognitive Structure, Theory and Analysis* (New York: Oxford University Press, 2006).

4. Anja Volk, W. Bas de Haas, and Peter van Kranenburg, "Towards Modelling Variation in Music as Foundation for Similarity," In *Proceedings of the 12th International conference of Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (Thessaloniki), 2012, 1085–1094.

5. Irene Deliége, "Similarity Relations in Listening to Music: How do they come into Play?," *Musicae Scientiae* Discussion Forum 4A (2007): 9–37; Leonard Meyer, *The Spheres of Music: A Gathering of Essays* (Chicago: The University of Chicago Press, 2000).

Most prominently studied is the apparent importance of melodic similarity in theme and variation sets. Robert Welker (1982) assumed that when measuring melodic similarity in a set of transformations on a theme, the theme would be the central tendency (i.e. the central or typical value of a probability distribution) of the transformations.[6] After listening to each of the five transformations, he asked participants to draw the melodic contour that best described all five pieces (the central tendency of them), which, according to Welker, should amount to the theme's melodic contour. The study showed that participants, both novices and experts, extracted the melodic contour of the theme, and identified it as the melodic similarity between the pieces. Iréne Deliége (2007) also explored melodic similarity between variations and their theme.[7] She took two themes (motives A and B) from the same variation set and asked her participants to identify the frequency of each motive in a piece. The participants were also asked to assess the degree of similarity between pairs of motives on a scale of 0 (no similarity) to 6 (total similarity). The study showed that the frequency of the different motives was rated correctly, and where motive B occurred more frequently, it received a higher grade than motive A, which was less prominent in the piece. Also, participants judged variations of motive B as more similar than variations of motive A, which Deliége attributed to the fact that variations of motive B retained the original cues of the motive.

Other research has used theme and variation sets to determine which 'structural features' are relevant to the perception of similarity. Stephen McAdams and Daniel Matzkin (2001) explored whether there was a difference in perceived similarity between a theme and its variation, dependent on where in the hierarchical structure changes exist (such as foreground vs background as proposed by Schenkerian theory).[8] They composed pieces of music where the surface characteristics were very similar to the theme but the underlying structures were different, and, by contrast, pieces that shared an underlying structure with the theme but had a variety of surface-level variations on the pitch content, rhythmic content, or both. Interestingly, participants rated those pieces that shared the same underlying structure as the theme as significantly more similar to the theme than those that did not. They also found that surface changes had different ratings depending on the theme that was subjected to the changes, suggesting that the context of the theme is fundamental in similarity judgements and that the nature of the theme was an important variable in considering these similarity judgements.

It is clear from previous research that the formal type of theme and variation lends itself to the study of audible perceptions of similarity. This chapter will examine which musical-theoretical approaches to similarity align with auditory perception. The listening study and analyses in this chapter use extracts from Johannes Brahms's *Studien für Pianoforte: Variationen über ein Thema von Paganini Op.35*, mostly written in

6. Robert Welker, "Abstraction of Themes from Melodic Variations," *Journal of Experimental Psychology Human Perception and Performance*, 1982, 435–447.

7. Deliége, "Similarity Relations in Listening to Music: How do they come into Play?"

8. McAdams and Matzkin, "Similarity, Invariance, and Musical Variation."

Vienna in the winter of 1862–63.[9] This set of piano variations, written for Carl Tausig who premiered the set in 1867, stands at the end of a line of large-scale piano variation sets that Brahms composed in the late 1850s and early 1860s. Carl Tausig, a pupil of Franz Liszt, was recognised as a successful virtuoso and this set of variations is often, therefore, seen as a challenge set by Brahms for Tausig and the Weimar school of which he was a part.[10] The theme has a very recognisable melody that has inspired over 20 other theme and variation sets; including Franz Liszt's *Études d'exécution transcendante d'aprés Paganini* for solo piano (1838), Sergi Rachmaninoff's *Rhapsody on a Theme of Paganini* Op. 43 (1934) for piano and orchestra, Witold Lutosławski's *Variations on a Theme by Paganini* (1940–41) for two pianos (and his 1978 version for piano and orchestra) and Andrew Lloyd Webber's *Variations* (1977) (an album originally for cello and rock band, used as the theme for ITV's The South Bank Show 1978–2010). The melody is from Niccoló Paganini's Caprice Op. 1, No. 24 in A minor (itself a theme and variation set) and may have been frequently adopted in part because it is simple and infectious, featuring a tight motivic structure and simplicity of harmony, which has the potential for display and elaboration. Brahms's variation set explores a wide range of techniques, including difficult leaps, double chords, and polyrhythmic combinations.[11] Each variation concentrates on a particular technical challenge and is approached almost like a finger exercise, as Brahms's choice of title — 'Studien fur Pianoforte', instead of 'Variationen' — implies.[12]

The remainder of this chapter will introduce an online listening experiment (Section 2.2), featuring extracts from Brahms's Paganini Variations (the name often used to refer to the collection) to explore the audible perception of music similarity from a variety of music-theoretical perspectives. The listening experiment used 16 extracts from the two books of variations; these extracts were chosen based on their varying levels of similarity and dissimilarity to the theme, judged by different music theory techniques, and current prominent non-theoretical feature extraction methods (which happen to be mostly from the MIR domain).

This chapter aims to explore theoretical versus non-theoretical approaches for each musical feature (where possible). Though these paradigms are the focus of this chapter, often, the approaches of theoretical versus non-theoretical analysis are tightly coupled with either score- or audio-based approaches. Music-theoretical approaches originated in the score domain, where most analytical methodologies require a score to complete

9. Hans Kann, 'Preface' to *Variationen über ein Thema von Paganini Opus 35* (G. Heile Verlag, 1985).
10. Julian Littlewood, *The Variations of Johannes Brahms* (Plumbago Books, 2004).
11. Kann, 'Preface' to *Variationen über ein Thema von Paganini Opus 35*.
12. Kann, 'Preface' to *Variationen über ein Thema von Paganini Opus 35*; Littlewood, *The Variations of Johannes Brahms*.

the analysis.[13] Due to the current limitations in computational music theory (discussed in Chapter 1, Section 1.3) very few, if any, comprehensive computational approaches to music theory exist, and even when they do they still require a musical score (for which available encoded scores further limit us). Non-theoretical approaches to music similarity, or feature extraction, have become prominent features of exploration in MIR due to the affordances they provide for music recommendation, and because they do not require a score for analysis. Therefore, even though the purpose of this chapter is not to compare audio- and score-based methods, but instead to explore music theory's influence on predicting the perception of similarity between two extracts, the two are highly intertwined and unavoidably blurred.

This chapter shall discuss each question's materials and results before culminating in a discussion of the study's overall findings. Each question will introduce the extracts chosen (e.g. Section 2.4.1) highlighting how a music theory and feature extraction technique would assess the similarity of these extracts. Following this, each question's results section (e.g. Section 2.4.2) discusses the findings of that question, and begins to highlight the relationship between auditory perceptions of similarity and music-theoretical definitions of similarity. Finally, Section 2.9 proposes further exploration of whether harmony, Riemannian theory and melody align with audible perceptions of similarity. As we will see, these music-theoretical approaches aligned between 57% and 61% with the participants' rankings, suggesting there may be a relationship between audible similarity and these music-theoretical approaches. From this study, I narrow my thesis's scope to focus on harmonic similarity and the importance of Riemannian theory in instances of disagreement between aural harmonic transcriptions.

2.2 Design

For each question, I predict the participants' rankings of extracts' similarity or dissimilarity to the theme according to different music-theoretical approaches, including formal analysis, Riemannian theory, and Schenkerian theory.[14] These extracts are then presented to the participants for them to rank according to their similarity to the theme, without knowledge of the methodology I used to predict their ranking.

This online study featured 'closed' questions; that is, where a participant chooses from a list of possible answers,[15] producing standardised quantitative data that can be statistically analysed. Music similarity studies including Daniel Müllensiefen and Kalus Frieler

13. In some cases, if the analyst has a particularly good ear/memory, specific theoretical analysis methods may not require a score — especially to analyse a single phrase. For example, it is technically possible to create a Riemannian analysis without a score. However, very few people have a good enough ear to complete this.

14. This study also examined tempo. However, as I found no significant results for this feature, and therefore I will not include it in this discussion.

15. Arlene Fink, *How to Ask Survey Questions*, 2nd (California: Sage Publications, Inc., 2003).

(2007), Stephen McAdams and Daniel Matzkin (2001), and Iréne Deliége (2007), often asked for a Likert-scale-based ranking system.[16] However, I opted for a ranking system, as this enables participants to judge both the similarity between each extract and the extracts' similarity to the theme. This approach is closer to the methodologies of Rainer Typke et al. (2005) and Rainer Typke, Frans Wiering and Remco C. Veltkamp (2007), who asked participants to rank multiple melodies against a single repeated theme (or as they refer to it a 'reference melody'), although, it is worth noting that these studies only rated melodic similarity and asked participants to rank them explicitly based on this feature.[17]

This study first required each participant to complete a mandatory consent and participant information form (refer to Appendix A for screenshots of the different pages of the online questionnaire). The first set of questions featured demographic questions, including age (in a closed question format), musical performance qualifications (closed question), musical academic qualifications (closed question), instrument (in specified response format, meaning a participant typed a written answer to the question), participation in ensembles (boolean answer format, a choice between yes and no) and listening habits (closed question). The main questionnaire featured six questions in the same format throughout, with ordinal responses, of which five are discussed here.[18] The original Question 2 has been omitted from the discussion in this chapter as the results were of no significance (the question based on tempo). Each question provides the theme (called the 'seed song' in the study, a concept derived from the MIR community) which was the same for all six questions, although some featured the full extract (Question 1 and 5), and some a short eight-bar extract (Questions 2–4). For each question, I used four different extracts (from four different variations) to remove any familiarity bias caused by varying levels of exposure to an extract. I created the audio file (mp3) of each extract using general MIDI (extracted from Sibelius notation input) with a bit rate of 128 kbps and a sample rate of 44.1 kHz.

For each question, the user ranked the extracts in order, from those which were most similar (1) to the theme, to those that were least similar (4/5); this gave an ordinal measure. I chose to use four extracts for each question with the theme, following the recommendations of Arlene Fink who states one should use no more than five alternatives in ranking questions in self-administered surveys, as participants will not be able to remember any more than this.[19] I designed each question to explore a different specific music-theoretical approach to similarity (see Table 2.1), by providing the participant

16. Daniel Müllensiefen and Kalus Frieler, "Modelling Experts' Notions of Melodic Similarity," *Musicae Scientiae* Discussion Forum 4A (2007): 183–210; McAdams and Matzkin, "Similarity, Invariance, and Musical Variation."; Deliége, "Similarity Relations in Listening to Music: How do they come into Play?"

17. Rainer Typke et al., "A Ground Truth for Half a Million Musical Incipits," *Journal of Digital Information Management* 3, no. 1 (2005): 34–39; Rainer Typke, Frans Wiering, and Remco C. Veltkamp, "Transportation Distances and Human Perception of Melodic Similarity," *Musicae Scientiae* 11, no. 1 (2007): 153–181.

18. Fink, *How to Ask Survey Questions*.

19. Ibid.

Question	Approaches used to determine similarity	Extracts used
1	*Formenhlre* & Form segmentation	I/4, I/7, II/1, II/9
2	Melodic Contour	I/6, I/9, I/13, II/5
3	Schenkerian Analysis	I/2, I/3, I/11, I/14
4	Riemannian Theory & Harmonic analysis	I/12, II/4, II/11, II/12
5	The extracts the participant ranked as most similar for Questions 1–5	Dependent on previous answers.

Table 2.1: An overview of each question, showing the specific music-theoretical approach chosen for that question, along with the extracts chosen for each question (based on their varying levels of similarity or dissimilarity to the theme based on the chosen method). The participant was not aware of the chosen music-theoretical approach for each question.

with extracts judged as similar or dissimilar to the theme according to the question's applicable music-theoretical approach. The final question, 5, asked the participant to rate the similarity between the excerpts they chose as the most similar in the previous four questions.

2.3 Participants

The 162 participants were all over the age of 18. Of these participants, 35.8% (58/162) had extensive musical training, having completed an undergraduate, master's, or doctoral degree in music. A further 14.8% (24/162) of the participants had completed a GCSE, AS or A-Level in music.[20] The level of practical musical expertise of the participants was also recorded: 31.5% (51/162) had an ABRSM Grade 8 instrument (or vocal) qualification or equivalent; the majority of the participants, 63.6% (103/162), had been part of an ensemble or group. The majority of participants listened to music for 10–15 hours per week (46/162); see Table 2.2. The participants were mostly between the ages of 25–34 (50/162); see Table 2.3.

Hours Listening Per Week	Frequency
None	1.2%
Less than 1	1.2%
1–3	17.3%
3–6	26%
6–10	11%
10–15	28.4%
15 Plus	14.8%

Table 2.2: The proportion of participants that fell within the different listening hour per week categories.

20. See Footnote 54 on page 10 for a definition of GCSE, AS and A-Level.

Age	Frequency
18–24	27%
25–34	31%
35–44	13%
45–54	13%
55–64	12%
65–74	3%
75-Plus	1%

Table 2.3: The proportion of participants that fell within the different age categories.

2.4 Question 1

I will discuss the materials used for each of the five questions and the results. I have chosen to structure this chapter in this manner to enable the reader to keep my analysis of each extract in mind while reading participants' results. Therefore, this chapter will be divided into five mini-studies. Each question's materials section will provide a discussion of the methodology used in that question to determine the chosen extract's similarity to the theme. The section will conclude with a predicted rank ordering of the extracts based on their similarity to the theme; these predictions are used to examine participants' rankings with different music-theoretical approaches in each question's results section.

2.4.1 Materials

The first question aimed to interrogate notions of musical form with regards to similarity. This question used the entire theme because we establish the musical form over the full length of a piece. I chose the following four variations for this question: Book I, Variation 4; Book I, Variation 7; Book II, Variation 1; and Book II, Variation 9 (which will be referred to as I/4, I/7, II/1 and II/9 respectively). To analyse the form of the chosen theme and the variations, I use William E. Caplin's theory of formal functions.[21] The theme (see Figure 2.1) in A minor is a 12–bar theme, divided into four + eight bars with both parts repeated, making it 24 bars in total length. As Caplin notes, identifying form by measure length alone tells us nothing about the content of the groups or how they relate to one another.[22] Therefore, I will also use letter-based labelling of sections according to their melodic content and specific formal functions in order to analyse the piece's form.

The first part of the theme, Figure 2.1, features a two-bar phrase that moves from the tonic to the dominant using a staccato, broken rhythm. This two-bar phrase is repeated (bars 3–4) with changes made to the fourth bar to create elongated falling

21. Caplin, *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart and Beethoven*.

22. Ibid., p.9.

Figure 2.1: The theme, from Brahms's Paganini Variations Op. 35, annotated with the Form AABB, i.e. ‖:A:‖:B:‖(Adapted by removing the appogiaturas from the right hand that begin each of the first four bars. The broken chords in the left hand in section B have been changed to static chords).

octaves (bar 4) to finish on the dominant as a half cadence (the third type of formal function). The second part features the same rhythmic pattern, but presents a closed 8-bar structure using a model (short idea) that forms the basis for a sequential treatment for bars 5–10 before a perfect authentic cadence. This second half finishes with a rising octave in the right hand. The bass doubles the whole theme, which features increased harmonic acceleration. The theme has an overall formal structure of AABB,[23] or one-part repeated, then a second part repeated (i.e. ‖:A:‖:B:‖) — also known as binary form. This measure grouping, and the use of thematic materials and cadential function will be observed in each theme to see which variation is most similar to the theme.

We can also analyse the form of the theme by using a segmentation tool (a non-music-theoretical approach). This tool observes sections of internal similarity, and distinguishes these sections of similarity from each other, thus 'segmenting' the piece. For this approach, I utilise the Segmenter vamp-plugin for Sonic Visualiser (as discussed in Chapter

23. The musical form can be determined using different musical features, including harmony and melody. For this study, the form was determined first by the bar lines of the piece — showing the division into sections. Then, if required, the melodic information was used to determine the label for each section.

FIGURE 2.2: The theme, from Brahms's Paganini Variations Op. 35, annotated with the form ABCBC extracted using the Segmenter algorithm hybrid approach —aligned to the score used in Figure 2.1.

1, Section 1.4.3). This vamp-plugin uses timbral features, harmonic features, and a hybrid of the two, to create a segmentation of the piece and assign a letter label. I then compare these segmentations to see if the formal structure of a variation is similar to the themes. The Segmenter algorithm has a variety of settings; for this analysis, I selected the Hybrid Constant-Q feature type as this uses both timbral and harmonic features for its segmentation. The other plugin parameters selected were: 10 segment types, a minimum segment duration of segments of four seconds, 26460 audio frames per block, and a window increment of 8820. These were chosen based on the guidance given by Mark Levy on how to implement 'Segmenter' as a vamp-plugin for Sonic Visualiser.[24]

Analysing the theme using the Segmenter algorithm hybrid constant-Q feature returns the segmentation of the piece seen in Figure 2.2. I have aligned this with the bar numbers from Figure 2.1. Figure 2.2 shows that the first eight bars form one section, Section A. The algorithm does not identify the internal repeat within these eight bars (i.e. 'A' instead of 'AA'). Following these eight bars, the hybrid approach distinguishes two sections of four bars repeated, labelled B and C. Overall, the full structure of this theme is ABCBC.

24. See section 8 of the Queen Mary Vamp Plugins page: https://vamp-plugins.org/plugin-doc/qm-vamp-plugins.html

The extracts that were chosen for comparison to the theme in this question feature varying levels of similarity based on their formal structures. For example, extract II/9, according to Caplin's approach, has the same form as the theme, with part one repeated, and then part two likewise repeated: AABB. Similarly, it is 12–bars, divided into four + eight bars with both parts repeated, making it 24 bars in total length. The first section similarly finishes on the dominant with a half cadence, moving harmonically in alternate bars. The second part features increased harmonic acceleration, and the octave rise in the final bar leaves the second section (B) with a perfect authentic cadence.

In contrast, the Segmenter algorithm extracts the form AB for extract II/9, with A being the first four bars repeated, and B being the following 16 bars (Figure 2.3 shows the segmentation). Similarly, suggesting that the form is in two sections 'A' and 'B', but the alternation of 'B' and 'C' found in the theme's segmentation analysis does not occur in this variations analysis.

Each of the other extracts chosen had forms that were slightly more removed from the theme. For example, I/7 features two distinct sections which, again, could be labelled A and B. However, section A is eight bars (not four bars repeated), and section B is eight bars repeated with an alternate final bar. Though this theme is still 24 bars in length, it does not feature this repetition with the half cadence in the middle. The final section (B) finishes with a perfect authentic cadence, like the theme. Again, similar to the theme, section B progresses and develops with more significant harmonic movement, retaining only a few melodic elements of A (Figure 2.4). Observing the results of the Segmenter algorithm, for extract I/7 the form ABCBC is extracted. Therefore, using the Segmenter algorithm, the form is judged as the same as the themes (ABCBC).

The third extract chosen was II/1, which also features two distinct sections: A and B (see the annotation below the stave Figure 2.7). Though there is a repetition of the first section, it is varied, and the right-hand and left-hand swap roles (both A sections are four bars in length, the same as the theme). The left-hand also has a slightly different descending broken octave pattern rather than the right-hand original block octaves; this leaves the form of AA′B, as both A sections feature the same melodic material with the hands reversing roles. The B section, like the theme and I/7, begins similarly to A but develops the melody further and for longer, also featuring increased harmonic movement. This variation also includes a key change finishing, with A major instead of A minor. For extract II/1, the segmenter approach did not find any similarity between the segments, producing the musical form ABCDE (as seen above the stave in Figure 2.7).

Finally, I/4 prominently features the technique of repetition, but, instead of having two distinct sections, it is based on repetitions and variations of the same melodic material (A), producing the form AA′A″. Therefore, this variation is divided into three parts instead of the four or two parts seen in the theme and in previously discussed variations.

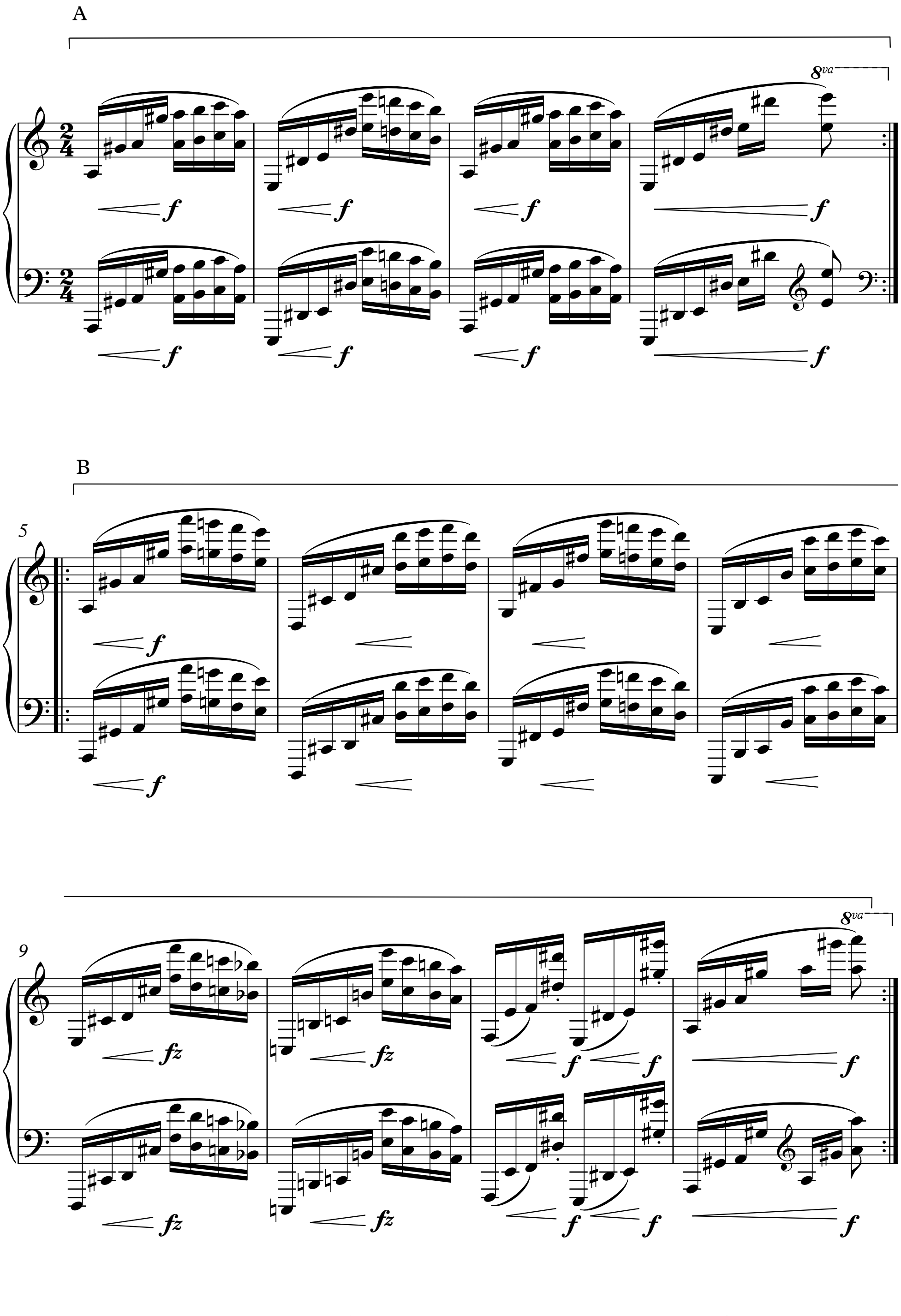


FIGURE 2.3: Book II, Variation 9 from Brahms's Paganini Variations Op. 35, annotated with the form AABB using traditional music-theoretical approaches. The segmenter algorithm extracted the form AB (without the section's internal repeats).

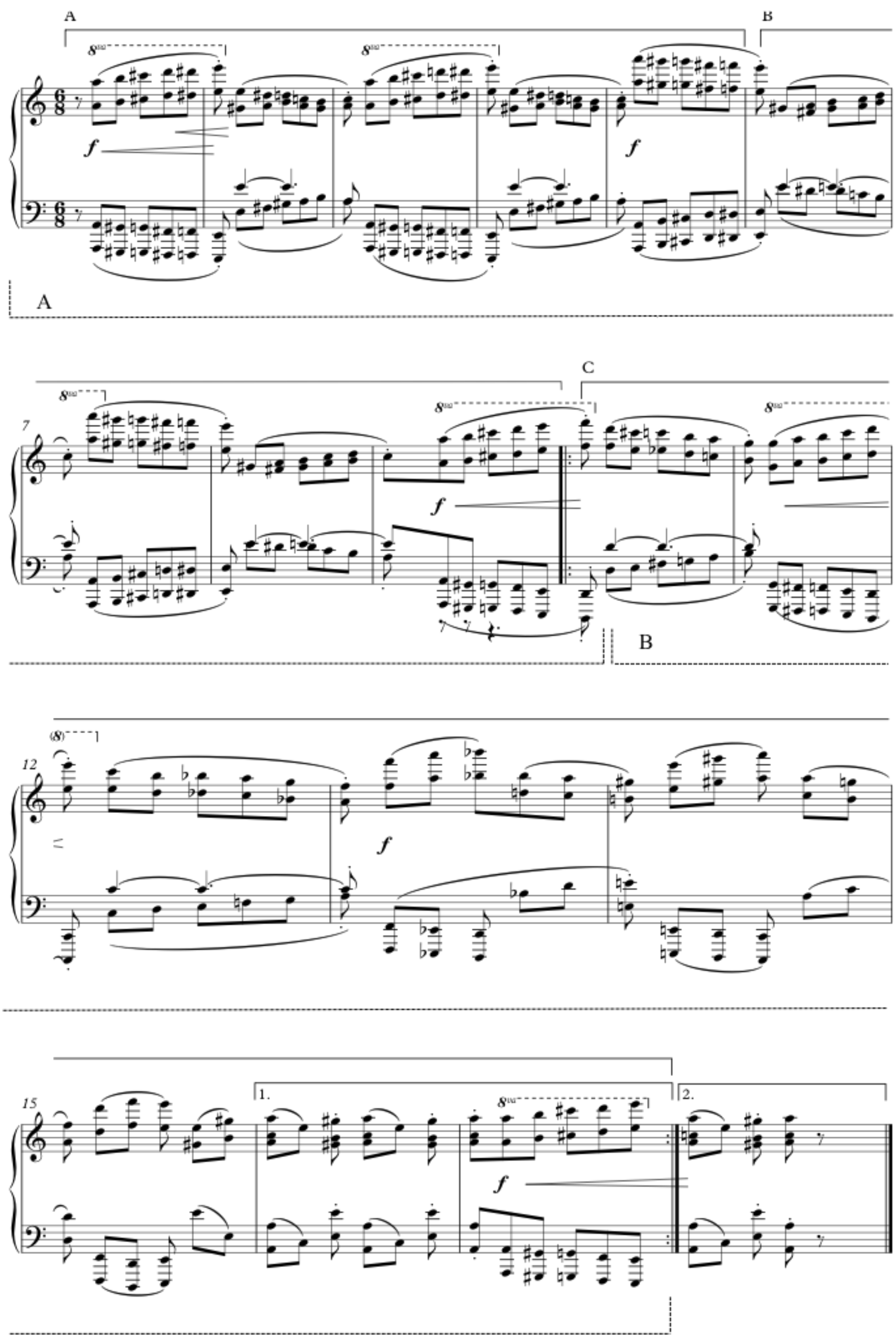


FIGURE 2.4: Book I, Variation 7 from Brahms's Paganini Variations Op. 35, annotated with the form ABB using traditional music-theoretical approaches (below stave), and the segmenter algorithm extracted the form ABCDC above the stave.

FIGURE 2.5: Book II, Variation 1 from Brahms's Paganini Variations Op. 35, annotated with the form determined by *Formenhlre* AA′B below the stave, and the form determined by the Segmenter algorithm (ABCDE) above.

FIGURE 2.6: Book II Variation 1 cont.

The variation finishes similarly to the theme with a perfect authentic cadence in A minor. The Segmenter algorithm does not seem to recognise the similarities between the segments; this could be because it is hard to perceive such details from the audio — or it may be that the algorithm does not perceive the nuanced similarities that we annotate with a prime symbol (′). Therefore, the approach identifies seven unique segments.

Overall, this makes the predictive rank ordering of all the experts using the theoretical approach taken from Caplin as (1) II/9 as the labelling of the formal structure is the same, AABB. Similarly, this variation features the half cadence to end the A sections, the perfect authentic cadence ending the B sections and the increased harmonic movement in the B sections; (2) I/7 with the form ABB. This variation features similar differences between the A section and B section in the theme, such as increased harmonic movement. The B section retains its repetition but with a slight variation to end the second time through, and closes both times with a perfect authentic cadence. The lack of four + four bars in section A made this extract less similar than II/9 to the theme. (3) II/1 with AA′B variation, similar to the theme, features two distinct sections of which the first has a repeat (though varied), but the B section is not repeated. Finally, (4) I/4 has

FIGURE 2.7: Book I, Variation 4 from Brahms's Paganini Variations Op. 35, annotated with the form determined by *Formenhlre* AA′A″ below the stave, and the form determined by the Segmenter algorithm (ABCDE) above.

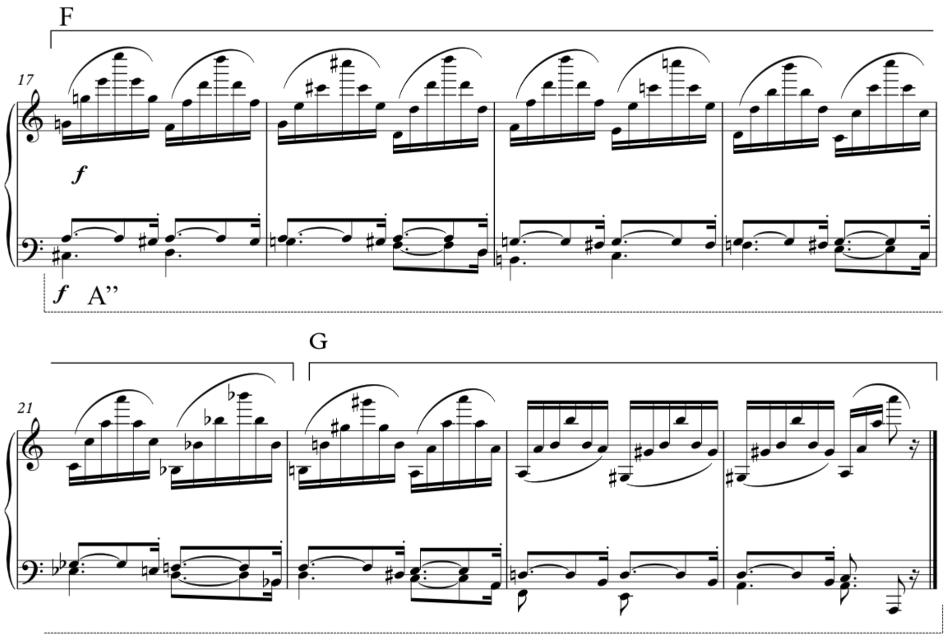


FIGURE 2.8: Book I variation 14 cont.

the formal structure AA′A″, and unlike the theme this variation has a single developing unit. However, this variation is broken into three sections instead of the theme's four.

For the Segmenter algorithm approach, I determined the predicted rank ordering by seeing which formal structure extracted was most similar to that of the theme (ABCBC). The predicted rank ordering is as follows: (1) I/7 as the extract with five sections one of them repeating like the theme (ABCDC); (2) II/9 with AB, and although only two distinct sections were determined, this is similar to the theme which could be seen as having two parts (A and BCBC); (3) II/1 with ABCD, and finally which features four distinct sections; (4) I/4 with ABCDEFG, with the highest number of distinct sections. Table 2.4 shows the overall prediction for each approach for Question 1.

Rank position	Music Theoretical	Non Theoretical
1	II/9	I/7
2	I/7	II/9
3	II/1	II/1
4	I/4	I/4

TABLE 2.4: The predicted rank order of similarity to the theme based on theoretical and non-theoretical approaches to form (Question 1), with 1 being the most similar to the theme, and 4 the least.

2.4.2 Results

For Question 1, there was a strong consensus between participants on the extracts most and least similar to the theme (seen in Table 2.5). Specifically, 62% (101/162) of the

Rank	I/4	I/7	II/1	II/9
1	28%	5%	5%	62%
2	36%	30%	11%	23%
3	22%	38%	29%	11%
4	14%	28%	55%	3%

TABLE 2.5: The proportion of participants that voted for each extract in each rank position, for Question 1. This table has highlighted the extract with the highest proportion of participants for each rank position (1–4). The total number of participants was 162.

participants ranked II/9 as the most similar to the theme, and 55% (89/162) of the participants ranked II/1 as the least similar to the theme (Figure 2.9). The consensus for the other rank positions was weaker, with I/4 having 36% agreement for rank position 2 and I/7 having 38% agreement for rank position 3. This consensus is also seen in the modes for each rank positions, confirming the most popular ordering of (1) II/9, (2) I/4, (3) I/7, and (4) II/1. This rank ordering does not align entirely with either predicted rank order, as seen in Table 2.4. However, Caplin's theory of formal functions 'correctly' predicted the variation in rank position 1. The rank order determined by the modes does not align with the predicted average results, where II/1 has a lower average rank position (2.57) than I/7 (2.62). For completeness, II/9 has an average rank position of 2.29; and I/4 an average rank position of 2.52 — leading to the average rank ordering (1) II/9, (2) I/4, (3) II/1, and (4) I/7, which aligns with rank positions 1 and 3 as predicted by my music-theoretical analysis.

There was no significant relationship ($t(161)=26.5$, $p=0.05$) between the predicted rank-ordering determined by the music-theoretical approach to form and the participants' rankings ($M=4.07$, $SD=1.96$) when using a one-tail t-test.[25] A lack of statistical significance means that the population does not align significantly with the predicted rank

25. The statistics in this sentence explore the significance of the relationship between music theory's prediction and participants rankings. Firstly, 't' is the test statistic value for the degrees of freedom (participants minus 1). The degrees of freedom for this statistical test was 161, and the test statistic is 26.5. The p-value states the significance level of 0.05, relating to 95% confidence level. The second set of statistics, the mean (M) and the standard deviation (SD), are used to calculate a one-tail t-test. The t-test observes the significance of the mean compared with a perfect alignment of no difference between the predicted rank ordering and participant's rank ordering. The standard deviation shows the spread of data. The mean of the absolute difference allows us to compare the results retrieved with those predicted, to observe whether participants ranked the extracts in a manner aligned with the music-theoretic or Segmenter algorithm methods.

Each participant's rank ordering is taken in order to see how far their ordering is from the predicted ordering — this figure is called the absolute difference. For example, II/9 was predicted in rank position 1 using the music-theoretical approach. If a participant placed II/9 in rank position 2, there is an absolute difference of 1 (absolute meaning that there are never minus numbers; it is the whole value distance). We take the sums of these absolute differences for each participant, which will return either 0 (all ranks are in the same order as the predicted order), 2, 4, 6 or 8 (ranking by participants is precisely the reverse of the predicted order). The mean of the absolute differences gives us a value for how close to the predicted rank order, or how close to 0, the participant's rank order is. For this example, the mean of 4.07 shows that about half the rank positions retained their order (0 being all rank positions retained, 8 being no rank positions retained ordering — the complete opposite). 4.07 is not a strong relationship between the population and music-theoretical approaches of measuring similarity in form, as this mean falls in the middle of the possible range.

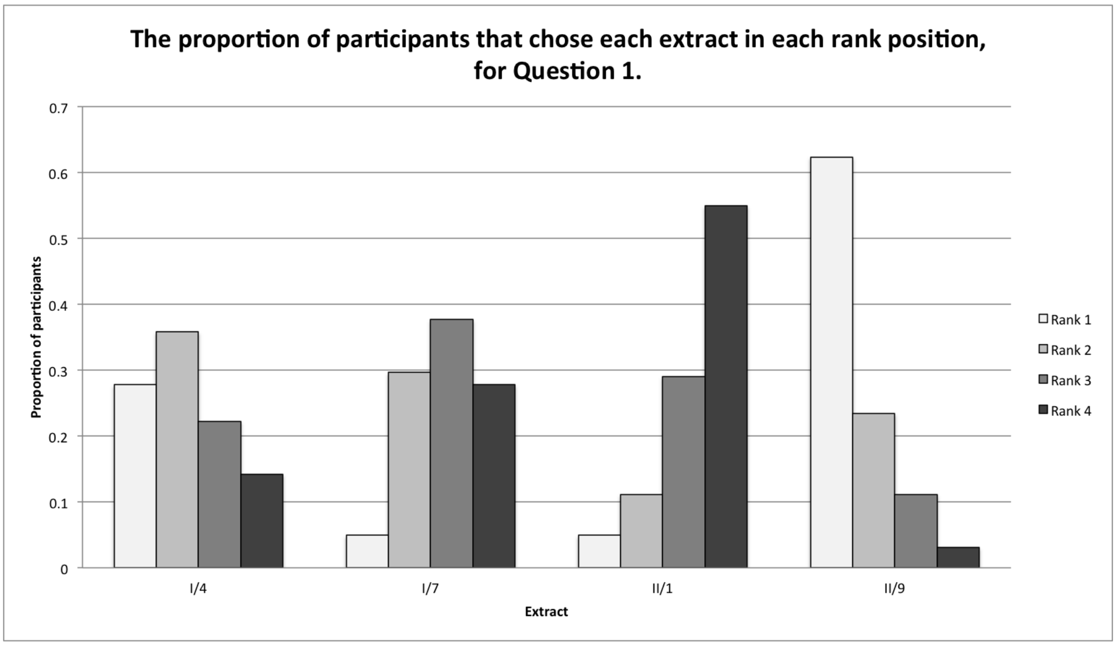


FIGURE 2.9: The proportion of participants that chose each extract in each rank position, for Question 1. The total number of participants was 162.

ordering. Form, as determined by the Segmenter algorithm, was less aligned with the population's ordering, having a higher mean (M= 5.22, SD= 1.62); this was also not statistically significant (t(161)=41.16, p=0.05). Table 2.4 and Table 2.5 show that neither approach aligned consistently with participants' rankings as the participants agreed only with the music-theoretical approach for extract II/9's position. Although, perhaps significantly, this was the rank position predicted by using Caplin's theoretical approach and judged by the participants to be the most similar extract to the theme.

I now turn to consider the effect of different demographics on participant rankings. For this question, ensemble experience was the only demographic feature to significantly impact the amount the participants' rankings aligned with my predictions. Table 2.6 shows the mean absolute differences for the demographic categories of having and not having ensemble experience compared with each of the predicted rank orderings. Figures 2.10 and 2.11 visualise the data from Table 2.6; the figures show that the participants in the category 'has ensemble experience' aligned with the music-theoretical approach (M=3.96, SD=1.96) and the Segmenter algorithm method of determining similarity (M=5.51, SD= 1.94) more than that of the category 'does not having ensemble experience' (theory being M = 4.31, SD =1.96; Segmenter being M=4.28, SD = 1.94). The results of the t-test analysis in both scenarios revealed the difference between the means of having and not having ensemble experience was significant (with theory t(161)= -0.98, p= 0.05; and Segmenter t(161)= 0.72, p= 0.05). Consequently, being part of an

Demographic	**Music-theoretical Mean Absolute Difference**	**Segmenter Mean Absolute Difference**
Has ensemble experience	3.96	5.15
Does not have ensemble experience	4.31	5.34

Table 2.6: The mean of the absolute differences between the music-theoretical and segmenter algorithm approaches predicted rank ordering, for Question 1, grouped by ensemble experience. There were 103 participants with ensemble experience and 59 without experience.

ensemble made a participant significantly more likely to rank the extracts in a manner that aligned with both the music-theoretical and Segmenter algorithm approaches. Therefore, in general, form plays a more significant role in judging similarity for those participants that had ensemble experience, although there might also be other factors involved.

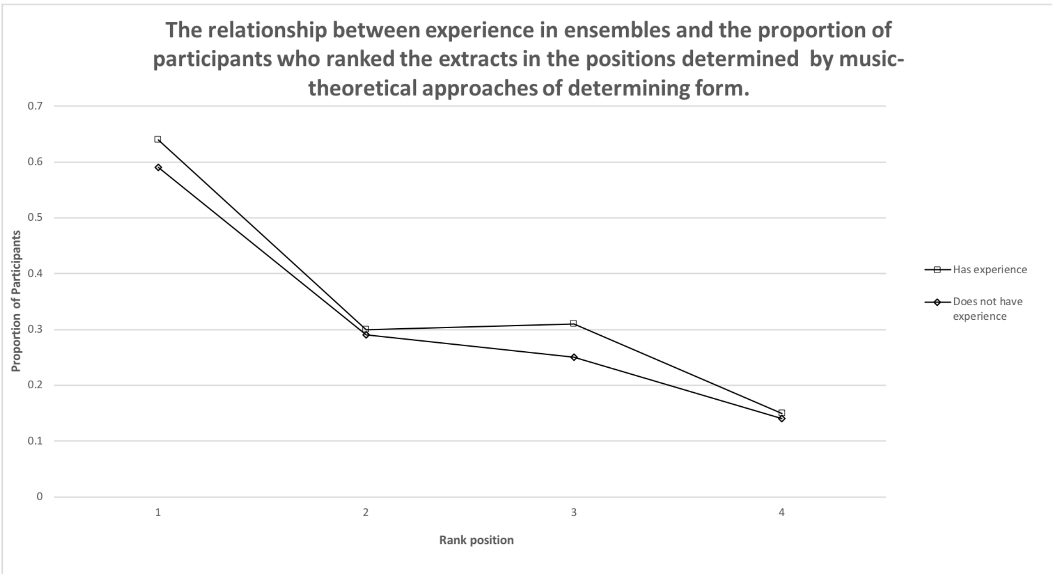


Figure 2.10: The proportion of participants that agreed with the rank order predicted by a music-theoretical approach for each extract in each rank position, grouped according to their ensemble experience, for Question 1. There were 103 participants with ensemble experience, and 59 without experience.

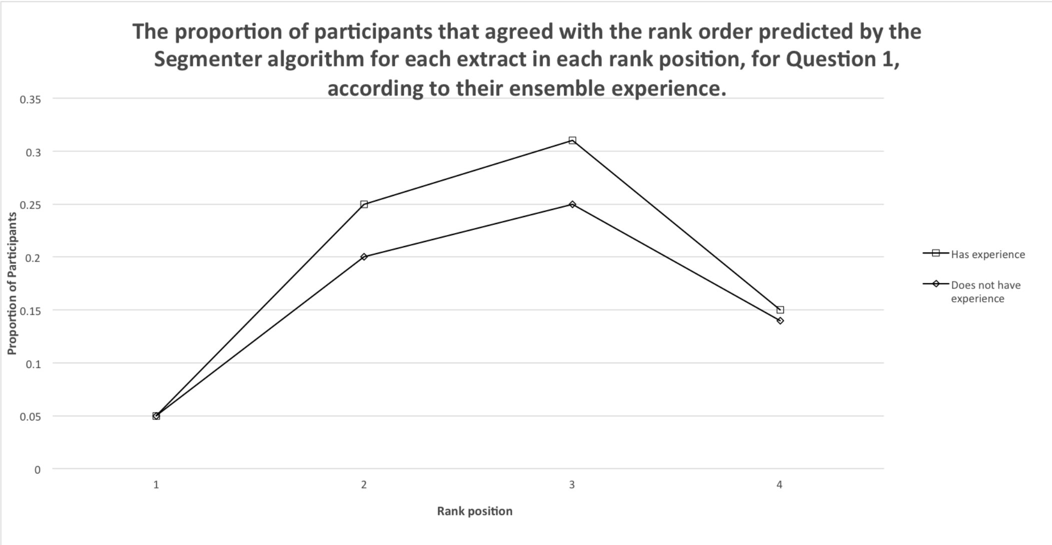


FIGURE 2.11: The proportion of participants that agreed with the rank order predicted by the Segmenter algorithm for each extract in each rank position, for Question 1, according to their ensemble experience. There were 103 participants with ensemble experience, and 59 without experience.

2.5 Question 2

2.5.1 Materials

Question 2 focuses on the theme's melody in the variations and assumes similarity based on the variations' closeness to the theme's melodic line. For this question, the first eight bars (four bars repeated) of the theme and each variation were used. The melody line in the theme is simple, mostly revolving around the notes a/e (i.e. $\hat{1}$ and $\hat{5}$) (see the first four bars of Figure 2.1). There is a rhythmic pattern of ♪𝄿𝅘𝅥𝅯♬♬ in the first three bars before being elongated by descending octaves in the final bar. The first three bars feature the starting note repeated three times, followed by a descending three-note figure starting from a third above the starting note. The first bar starts on an A, the second on an E, and the third again on an A. One prominent feature of this melody, therefore, is this alternating a^1, e^2/^{1}, a^1, e^2/^{1} ($\hat{1}$, $\hat{5}$, $\hat{1}$, $\hat{5}$) pattern, along with the descending three-note figure. To determine similarity to the theme, I looked for these key features of the theme (the alternating a/e pattern and the descending three-note figure) along the melodic contour rising and falling like the theme. For this question, I chose extracts from variations I/6, I/9, I/13 and II/5.

The melody of I/6 has a weak relationship to that of the theme, involving rhythmic augmentation of the third-leap and three-note descending figure. The theme's alternation of $\hat{1}$ and $\hat{5}$ (a/e) is seen in the harmonic alternation of `i` and `V` of I/6. This harmonic alternation reminds us of the alternating a/e figure from the theme; although, in this variation, the interchange is not in the melody line (see Figure 2.12).

I/9 was chosen as it is even more removed from the melody line than I/6. There is no surface-level melodic similarity (this includes the musical features and melodic contour); however, emphasis on the note A is present in the right-hand scales. Unlike the theme, this is not the alternation of $\hat{1}$ and $\hat{5}$ (a/e). Also, there is no apparent harmonic similarity, with the whole of I/9 featuring a rising chromatic scale, and little tonal centricity (see Figure 2.12).

In contrast, I/13 has a strong similarity to the melody line of the theme. The melody line adds further details to the part, with decoration between the notes, almost 'filling it in'. The original melody line is still prominent (featuring the alternation of the a/e), though the extract does not feature the falling semiquaver figures that leap up a minor third or the prominent rhythmic pattern. The piece does, however, retain some of the melodic notes as annotated in Figure 2.12.

Finally, II/5 again has a strong similarity to the theme's melodic line, featuring a developed version of the theme's melody (see Figure 2.12). II/5 takes the descending 3-note figure (which ends bars 1, 2, and 3 of the theme) and places it as a counter melody at the beginning of each bar. The theme's three repeated starting notes also feature in a higher octave elongated throughout the bar. The rests are still present, giving a similar rhythmic feel to the repeated notes (i.e. $\frac{2}{4}$ 𝅘𝅥𝅮𝄿𝅘𝅥𝅯𝅘𝅥𝅯 of the theme becomes $\frac{3}{8}$ 𝅘𝅥𝅮𝄿𝅘𝅥𝅯𝅘𝅥𝅮 in II/5). This melody is a development of the original melody; indeed, it is an excellent example of Arnold Schoenberg's concept of 'developing variation'.[26]

In participants' responses to this question, I assume that II/5 or I/13 will be chosen as the most similar due to their closeness to the original melody, since it is a development of the theme. Both have significant similarities, and either could be perceived as most similar depending on whether similarity is interpreted as the melodies sharing a greater number of similar components (in which case II/5 is most similar) or by the ordering of the notes (I/13 would then be most similar). As discussed previously, I have determined melodic similarity based on the variation featuring the prominent motifs of the melody (the descending three-note figure, and a/e alternation and the melodic rhythm). My predicted rank ordering for this question is, therefore: (1) II/5 with the most number of components from the themes melody; (2) I/13 due to sharing the next most similar melody (it does not feature the falling semiquaver figure) but retains a lot of the melody notes of the theme; (3) I/6 as it retains the a/e alternation; and (4) I/9 as it does not retain any of the melodic features. Table 2.7 shows the predicted rank ordering of extracts.

26. Schoenberg is famous for conceiving the concept of the 'developing variation', a formal technique where the concepts of development and variation are united. Schoenberg insists that the best way to produce a variation is by developing already existing material from the set.

FIGURE 2.12: The theme and the extracts chosen for Question 3; I/6, I/9, I/13, and II/5, from Brahms's Paganini Variations Op. 35.

Rank position	Music Theoretical
1	II/5
2	I/13
3	I/6
4	I/9

Table 2.7: The predicted rank order of similarity to the theme based on melodic similarity (Question 2), with 1 being the most similar to the theme, and 4 the least.

2.5.2 Results

In Question 2, the participants agreed upon the most and least similar extracts (with II/5 having 52% (84/162) agreement as the most similar, and I/9 having 73% (119/162) of participants stating that it is the least similar). Table 2.8 also shows a broad consensus of extract I/6 in rank position 2 with 45% agreement (73/162), and I/3 for rank position 3 with 46% agreement (74/162); similarly, the mode extract for each rank position confirms this ordering. The averages were as follows: II/5 = 1.83; I/6 = 2.07; I/13 = 2.47; and I/9 = 3.62. Therefore, there is a consensus for the ordering: (1) II/5, (2) I/6, (3) I/13 and (4) I/9 (as shown in Figure 2.13). There was no significant relationship ($t(161)=21.46$, $p=0.05$) between the predicted rank-ordering determined by melody and the general population ($M=3.41$, $SD=2.02$) when using a one-tail t-test. However, the predicted rank-ordering agrees with the participants in rank position 1 and 4, highlighting our ability to predict the most and least similar extracts.

The demographics of ensemble experience and age have statistically significant effects on the results. Table 2.9 shows the mean absolute differences between each ensemble category and the predicted rank-ordering determined by the melodic line. Figure 2.14 provides a visualisation of the proportion of participants that agreed with the extract predicted in each rank position. Category 'has ensemble experience' aligns most closely with determining similarity from the melody in rank position 1 and 4, and 'does not have ensemble experience' aligns most closely in the remaining two. The mean absolute differences show that having ensemble experience aligns the participant more closely ($M=2.95$, $SD=1.96$) with methods of determining similarity using melody than that of not having ensemble experience ($M=3.41$, $SD=2.14$). The results of the t-test analysis

Rank	I/13	I/6	I/9	II/5
1	20%	27%	2%	52%
2	24%	45%	7%	23%
3	46%	23%	17%	14%
4	10%	6%	73%	10%

Table 2.8: The proportion of participants that voted for each extract in each rank position, for Question 2. This table has highlighted the extract with the highest proportion of the participants for each rank position (1–4). The total number of participants was 162.

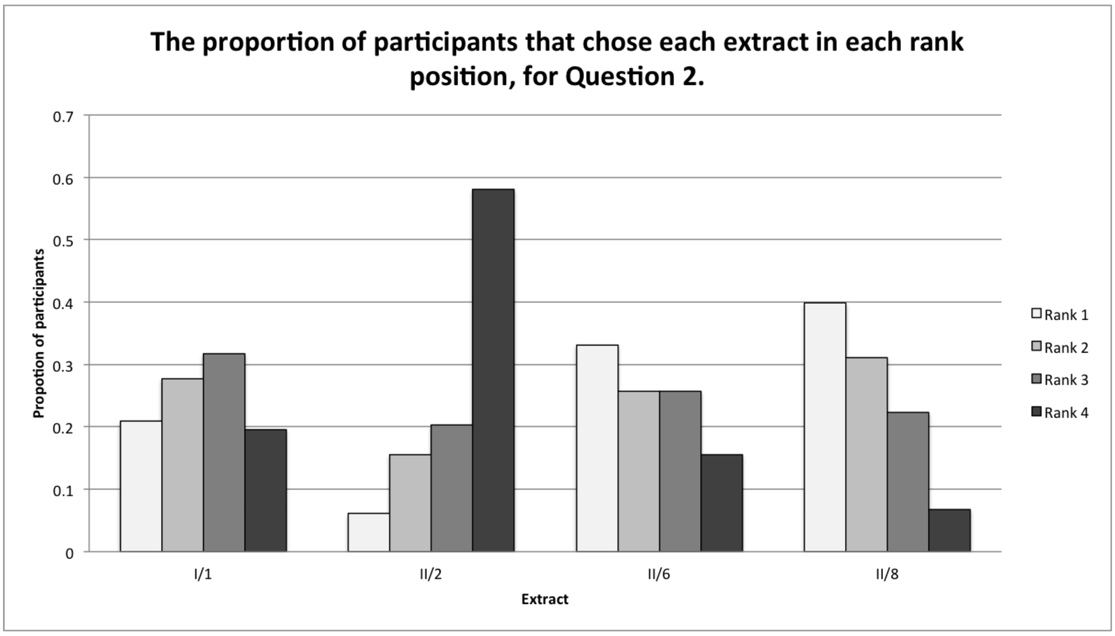


FIGURE 2.13: The proportion of participants that chose each extract in each rank position, for Question 2. The total number of participants was 162.

Demographic	Mean Absolute Difference
Has ensemble experience	2.95
Does not have ensemble experience	3.41

TABLE 2.9: The mean of the absolute differences between the melodic predicted rank ordering, for Question 2, grouped by ensemble experience. There were 103 participants with ensemble experience and 59 without.

revealed the difference between the means of having and not having ensemble experience was significant (with score t(161)=-1.39, p=0.05). Having ensemble experience made a participant significantly more likely to rank the extracts in a manner more closely aligned with methods of determining similarity from melody.

Figure 2.15 shows the relationship between age and the proportion of participants who ranked extracts in the positions predicted by my analysis in Section 2.5.1. The age category 25–34 sits at the top of the first half of Figure 2.15, overtaken by 35–44 in rank position 3 and 4. Table 2.10 shows that the category 25–34 has the lowest mean absolute difference (M=2.94, SD=2.53), and is the most aligned with the predictions made from melodic similarity. An ANOVA test — the analysis of variance that tests the difference between two or more means — reveals no significant difference between the groups at the 95% confidence level (F(5, 24)=2.53, p=0.057).[27] However, at the 90% confidence

27. F is a type of statistical test that an ANOVA calculates (the brackets, again, show the degrees of freedom, like the t-test). Unlike the t-test, the first number (5) shows the number of demographic

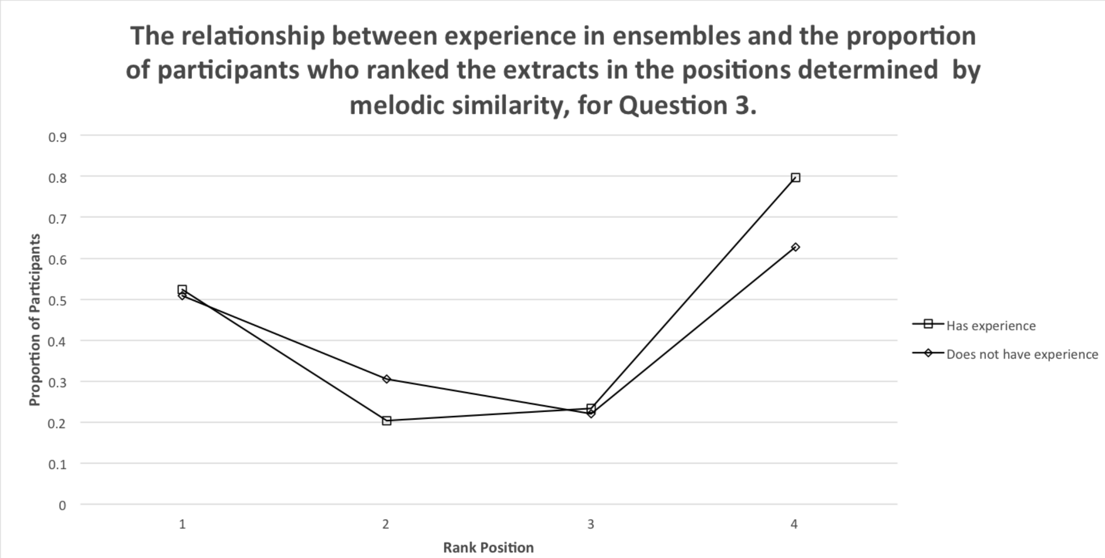


FIGURE 2.14: The proportion of participants that agreed with the rank order predicted by melodic analysis for each extract in each rank position, for Question 2, grouped according to their ensemble experience. There were 103 participants with ensemble experience and 59 without experience.

Demographic	**Mean Absolute Difference**
18–24	3.29
25–34	2.94
35–44	3.33
45–54	3.33
55–64	3.125
65 Plus	3.33

TABLE 2.10: The mean of the absolute differences between the melodic analysis approach predicted rank ordering and the populations rankings, for Question 2, grouped by age. For the different categories there were; 43 participants in 18–24, 50 in 25–34, 21 in 35–44, 21 in 45–54, 19 in 55–64 and 8 in 65 plus.

level, there was a significant difference between the means of the different age groups, as $0.057<0.1$. Nevertheless, observing the mean absolute differences (Table 2.10) for each age category, there does not appear to be a pattern as to whether being older or younger makes a participant more likely to rank the extracts in a manner more closely aligned with the predicted rank-ordering determined by the approaches to melodic similarity detailed in Section 2.5.1.

categories there are for the question, minus 1. The second number (24) is the residual degrees of freedom: the total number of instances minus 1 — in this case, 24 because there are five values for each demographic category (25).

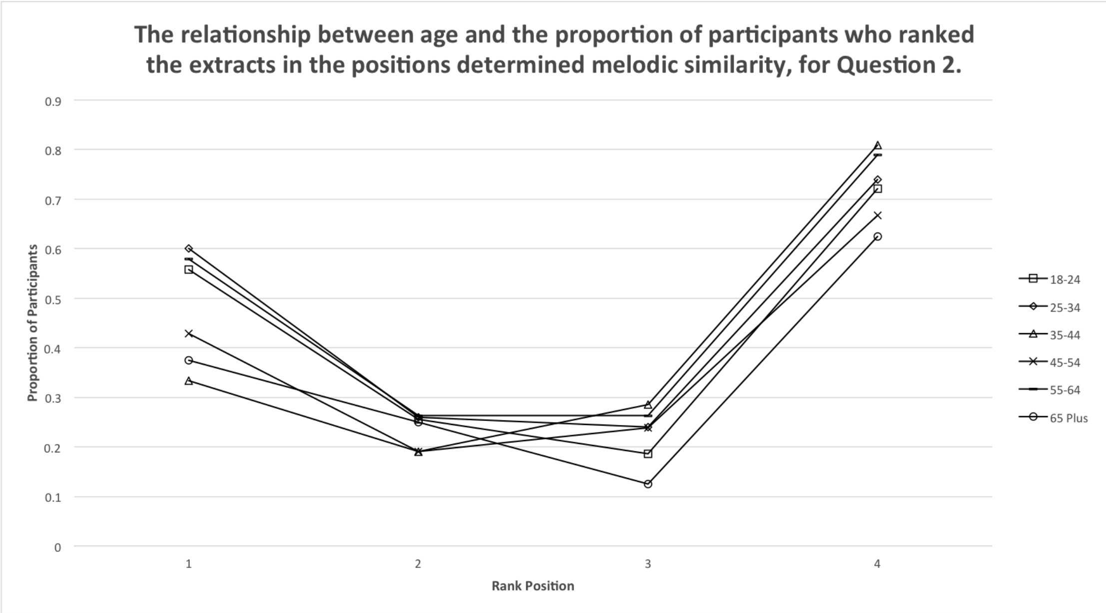


FIGURE 2.15: The proportion of participants that agreed with the rank order predicted by melodic analysis for each extract in each rank position, for Question 2, grouped according to their age. For the different categories there were: 43 participants in 18–24, 50 in 25–34, 21 in 35–44, 21 in 45–54, 19 in 55–64 and eight in 65 plus.

2.6 Question 3

2.6.1 Materials

Question 3 focuses on aspects of Schenkerian voice leading as a way of predicting music similarity.[28] In contrast to the last question, which considered melodic contour, this question considers similarity from a more systematic Schenkerian analysis, which considers the hierarchical importance of melodic notes instead of the foreground melodic contour. If we look at Heinrich Schenker's graph of the theme from Paganini's Op. 1 No. 24 (Figure 2.16), Schenker reduces the first eight bars (4 bars, repeated — up to the repeat mark seen in Figure 2.16) to a prolongation of the note E ($\hat{5}$), including an initial arpeggiation to the *Kopfton*. I, therefore, looked for this prolongation of $\hat{5}$ (or *Kopfton*–emphasis on $\hat{5}$) in the first eight bars of each variation to determine its similarity to the theme. For this question, I used the first eight bars (or four bars repeated) of the following variations: I/2, I/3, I/11 and I/14.

28. The reader is referred back to Chapter 1, Section 1.3.4 for a discussion on Schenkerian analysis.

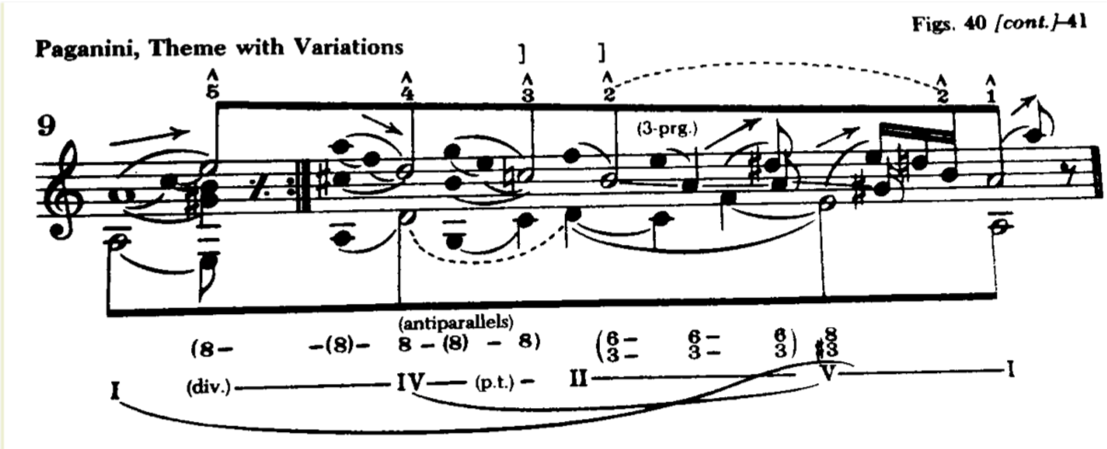


FIGURE 2.16: Schenker's Sketch of Paganini's Theme from Op. 1, No. 24 from Heinrich Schenker, *Free Composition* (*Der freie Satz*), vol. 3, book. 2 of *New Musical Theories and Fantasies*, ed. and trans. Ernst Oster (Hillsdale, NY: Pendragon Press, 1977), Fig. 40/9.

The foreground and background reductions of the extract taken from I/2 (see Figure 2.17) highlight a slight voice-leading emphasis of the note E. In the foreground structure, almost every bar features a prolongation of the note E by a neighbour note. The prominence of E ($\hat{5}$) in this variation is present through the alternation between E and another note (B, F or D♯) in the foreground, which features in one hand throughout the variation (starting in the right hand and moving to the left hand in bar 6, returning to the right hand in bar 7, and then back to the left hand in bar 8). The 'other' hand (the hand not featuring the prominence of E) features very little of this pitch class. The middleground has a prominent E, as the music starts with an e^4 rising to e^5, passing through b^4 before returning to e^5. The middleground, therefore, highlights the prominence of the pitch class 'E' throughout this extract; although, there is still a movement away from and towards the pitch class, elaborating the prolongation. The harmony, like the theme, features a tonic/dominant alternation; because these two chords feature (in their inversion) the pitch class E, its prolongation is also present in the harmony. Therefore, there is an emphasis on E in this variation, but it is not the strong rise to *Kopfton* (note E ($\hat{5}$)), like that observed in the theme.

I/3, like I/2, also features the pitch class E, with an emphasis on the pitch class as the monophonic line rises to and descends from the pitch throughout the first eight bars (see Figure 2.18). The harmony still emphasises the pitch class through a tonic-dominant alternation with more emphasis placed on the dominant: the dominant lasts two full bars, whereas the tonic only features in the last few notes of the bar. This extract

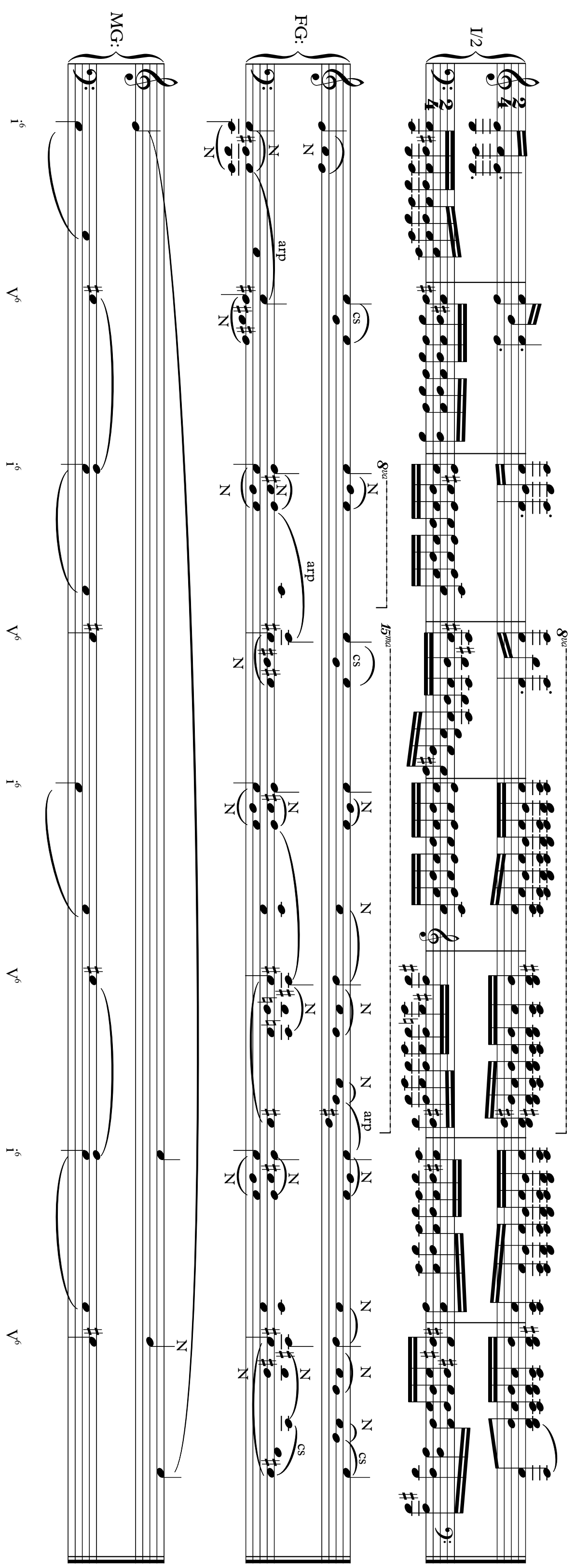


FIGURE 2.17: Schenkerian foreground and middleground reduction of Book I, Variation 2 from Brahms's Paganini variations Op. 35.

reduces to a prolongation of the note E in the right hand, with the left hand alternating between i and V — making it more similar to the theme than extract I/2.

In contrast, I/11 places less emphasis on the pitch class E (see Figure 2.20). The note is present in alternating bars and reflects the original harmonic movement (but transposed to A major), alternating I–V throughout. This alternation is present in both parts and reflected in the middleground structure, where the chords alternate. The pitch class E is most prominent in the foreground in the neighbour-note figure of E–D♯–E. The variation features a greater prominence of the pitch class A in the bottom voice of both hands in an alternating figure. For Question 3, this extract is the least similar to the theme.

If we observe just the notes of extract I/14, it appears on the surface to have the highest prominence of the pitch class E, and thus be the most similar to the theme. The first four bars (see Figure 2.21), feature the right-hand middle voice playing a repeated E (alternating with D♯) throughout, along with the pitch class featuring in the other parts as well. In bars 5–8, the E moves into the left-hand but still alternates with a D♯ every other note in the middle voice. The Schenkerian reductions (foreground and middleground shown in Figure 2.21) show a strong presence of the pitch class E, with one part featuring a prolongation of the note at all times. The harmony of extract I/14 progresses and moves through a `i–V` progression. As this extract has an elongation of the pitch class E throughout (in at least one part), I identify it as the most similar to the theme.

Using aspects of Schenkerian voice-leading, the predicted ordering of these extracts compared to the theme is: (1) I/14 as it has the highest prominence of the pitch class E; (2) I/2 as it features some prolongation of the note E alternated with another pitch class; (3) I/3 as it still has some prolongation of E, but, there is an equal emphasis on the dominant A; and (4) I/11 which features more of a prominence of the note A than E (as showing in Table 2.11).

Rank position	**Music Theoretical**
1	I/14
2	I/2
3	I/3
4	I/11

Table 2.11: The predicted rank order of similarity to the theme based on Schenkerian analysis (Question 3), with 1 being the most similar to the theme, and 4 the least.

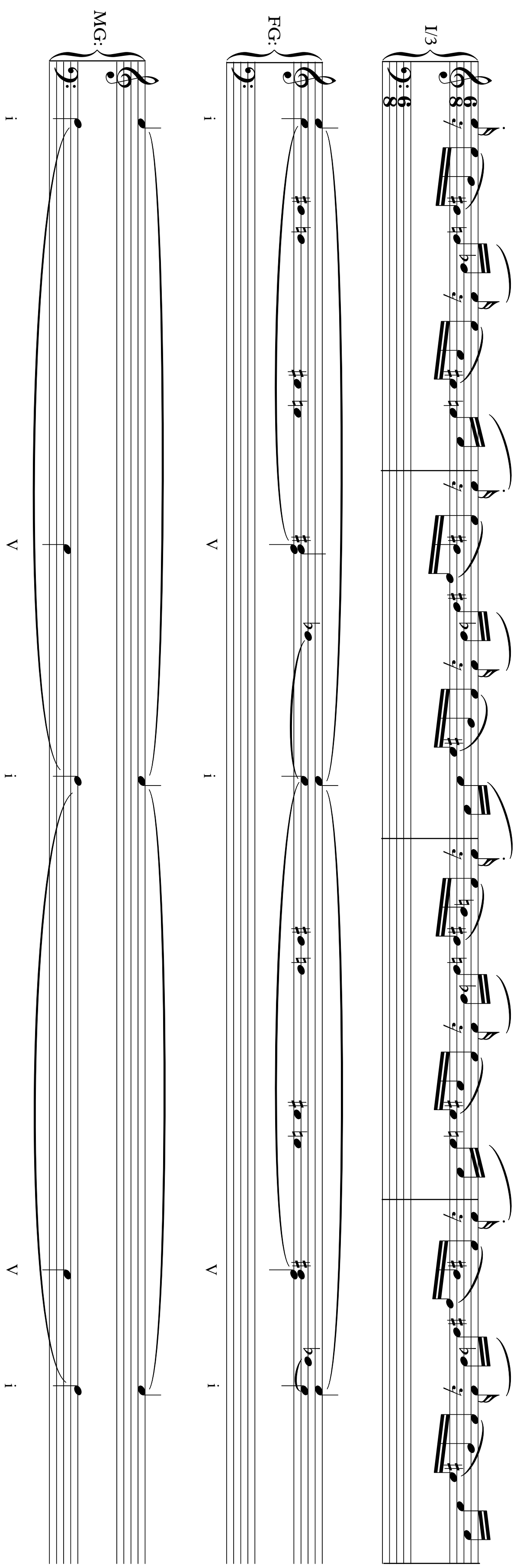


FIGURE 2.18: Schenkerian foreground and middleground reduction of Book I, Variation 3 from Brahms's Paganini variations Op. 35.

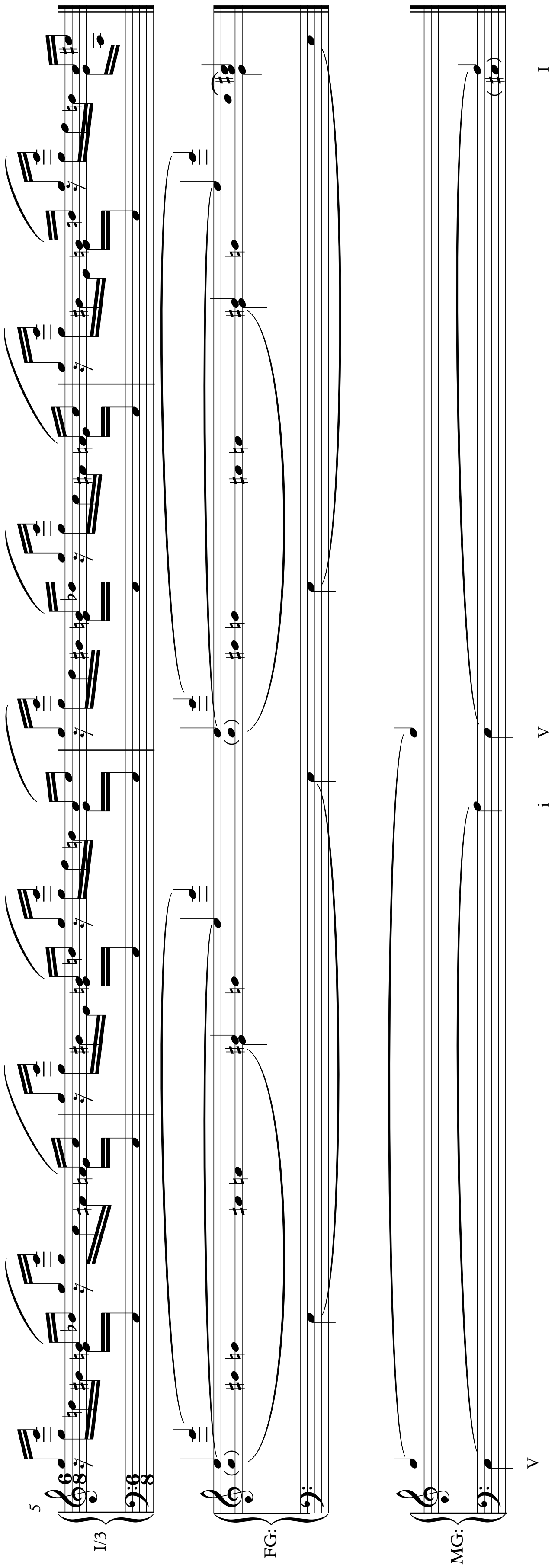


Figure 2.19: Book I Variation 3 cont.

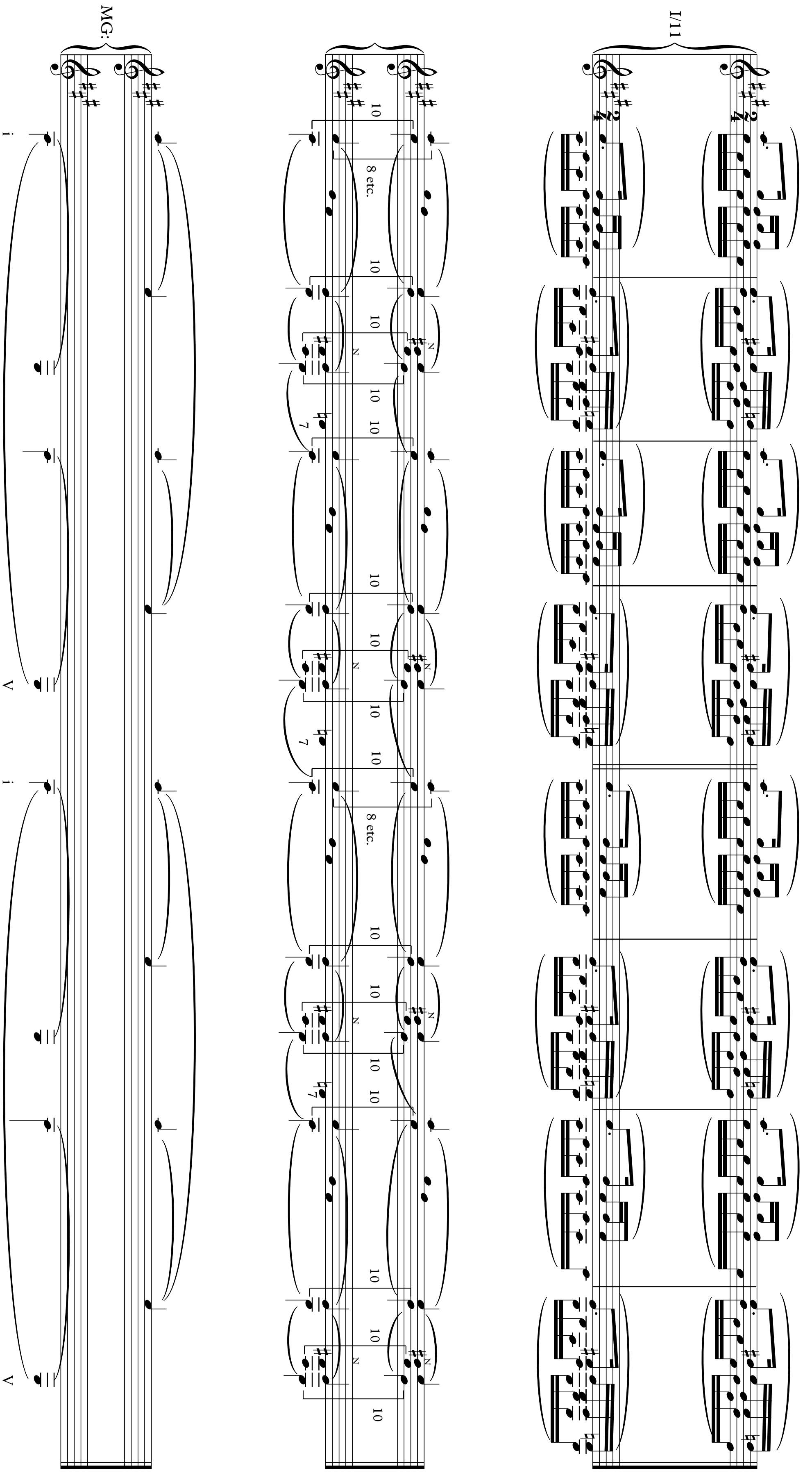


FIGURE 2.20: Schenkerian foreground and middleground reduction of Book I, Variation 11 from Brahms's Paganini variations Op. 35.

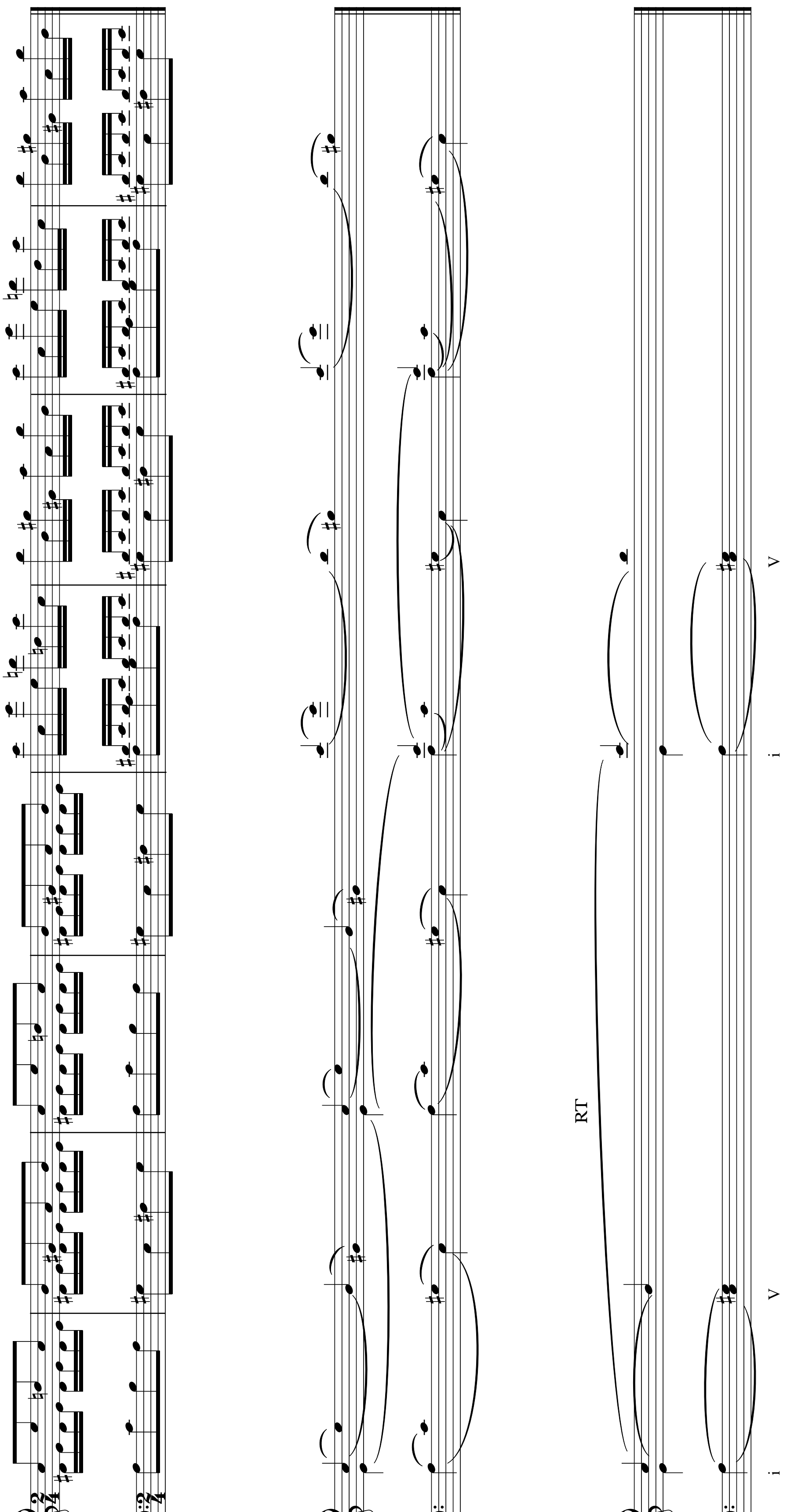


FIGURE 2.21: Schenkerian foreground and middleground reduction of the first eight bars of Book I, Variation 14, from Brahms's Paganini Variations Op.35.

2.6.2 Results

There was a relatively strong consensus for rank positions 1 and 4. I/14 has 37% (60/162) agreement for rank position 1, and I/11 has 49% agreement for rank position 4 (79/162). This consensus is not as strong as for the previous questions, as seen in Figure 2.22 and Table 2.12. The average rank-ordering, based on the average rank position of each extract, is as follows: (1) I/14 (an average of 2.1); (2) I/3 (an average of 2.22); (3) I/2 (an average of 2.52); and (4) I/11 (an average of 3.15). There was no significant relationship (t(161)=48.86, p=0.05) between the predicted rank-ordering determined by Schenkerian analysis (Table 2.11) and the general population (M=6.23, SD=1.6) using a one-tail t-test. However, as we have seen previously, the predictions made by music theory as to the extracts most and least similar align with the participants' most agreed extracts for those rank positions. No demographic features produced a significant difference between the demographic groups for this question.

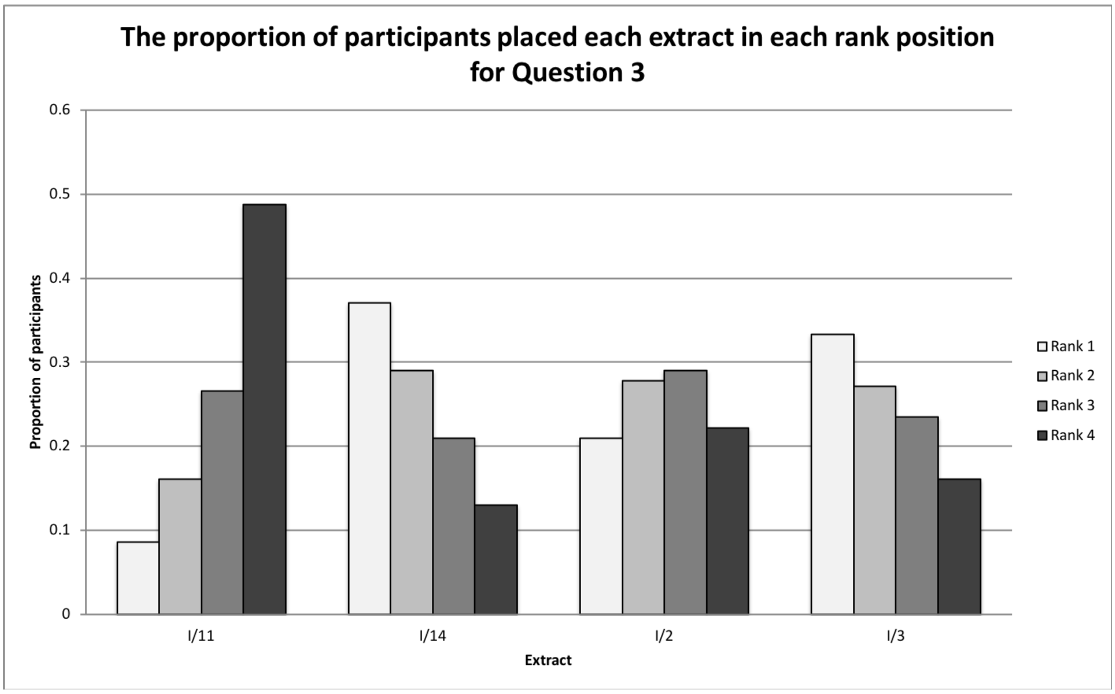


FIGURE 2.22: The proportion of participants that chose each extract in each rank position, for Question 3. The total number of participants was 162.

Rank	I/11	I/14	I/2	I/3
1	9%	37%	21%	33%
2	16%	29%	28%	27%
3	27%	21%	29%	23%
4	49%	13%	22%	16%

TABLE 2.12: The proportion of participants that voted for each extract in each rank position, for Question 3. The extract with the highest proportion of the participants for each rank position (1–4) is highlighted. The total number of participants was 162.

2.7 Question 4

2.7.1 Materials

The final question to feature new musical material, Question 4, looked at similarity through the lens of Riemann's theory of harmonic functions and harmonic similarity through feature extraction.[29] The theme's first four bars (eight in total with the repeat, Figure 2.23) feature a distinct tonic/dominant alternation in the harmony, as already discussed in Section 2.6 concerning Schenkerian prolongation. This question examines how far the harmony of each extract strays from this tonic/dominant alternation and I used this to predict the similarity of the extracts. For this question, the variations chosen were I/12, II/4, II/11 and II/12.

Firstly, according to Riemannian theory, the first four bar extract, taken from the opening of I/12 (Figure 2.23), has a much more frequently changing harmony than the theme, producing the progression `T–Tp–T–t–Tp–Dl–D` (my Riemannian 'primer' can be found in Chapter 1 Section 1.3.1). Figure 2.23 compares the Riemannian chord substitutions used in the selected variations with those used in the theme. In I/12, bar 1 first features a `T` instead of the `t` present in the theme, meaning that one substitution (the *Variante*) has been applied. Later in the bar there is the introduction of a `Tp` instead of the `t` in the theme, applying two substitutions: the *Variante* to take us from `t` to `T`, and the *Parallele* to take us to `Tp`. The bar concludes with one more harmonic change, returning to `T`, which again is one substitution (the *Variante*) from the `t` of the theme. The second bar begins with the tonic minor (`t`). This change does not retain the harmonic function. Since, in this chapter, I use the dominant transformation defined by neo-Riemannian theory to enable us to count the harmonic differences, the function is also changed (neo-Riemannian defines a set of further transformations, such as the Dominant which moves a chord to its Dominant, see Chapter 1 Section 1.3.2 for more detail). This means the `D` in bar 2 of the theme is transformed by two substitutions `D–>d–>t`, again utilising the *Variante* substitution, and then the dominant transformation. The harmony then moves to `Tp` roughly halfway through the bar, utilising four substitutions. The end of bar 2 replaces the `D` from the theme with `Dl`, a substitution of one (the *Leittonswechsel*).

29. I refer the reader Chapter 1 Section 1.3.1 for an explanation of Riemann's theory of harmonic functions

Bar 3 has the harmony `Dl`, which is three substitutions (again changing the function of the theme from tonic to dominant, `t–>d–>D–>Dl`), and then the harmonic change of `D`, which is two substitutions (`t–>d–>D`). The final bar sees the harmony's dominant function retained, meaning there are 0 substitutions for this bar. If we add together all the substitutions applied to change the harmony from the theme into I/12, we get a total of 15.

II/4 (see Figure 2.23) features fewer changes to the harmony. The tonic major substitutes the tonic minor (`t` to `T`), which is one substitution for each of bars 1, 3, 5, and 7. There are then changes made in the second and sixth bars to the dominant, moving it to $\mathtt{D}^{(\mathtt{rootless\ +\ 7})}$. I count this as a single substitution from the Dominant present in the theme; we have altered the chord removing the root and adding the 7$^{\text{th}}$. There are a total of six substitutions.

The next extract, II/11, has no changes in the harmony (see Figure 2.23). Since it preserves the original tonic/dominant alternation, the harmony is the same, and there are no substitutions.

The final extract, II/12, is very similar in its pattern of substitutions to the theme; however, the variation is originally in the key of F major. This variation is transposed to A major so as not to have the key distance affect participants' judgement of similarity. The harmonic progression is `T–D–Tp–Dp–DP`, resulting in six substitutions to the harmony.

Overall, by measuring similarity using Riemannian theory, the ranking order is predicted as: (1) II/11 because it has zero substitutions from the theme; (2) II/4 because it has six substitutions from the theme; (3) II/12 because it has six substitutions from the theme; and (4) I/12 because it has 15 substitutions from the theme. II/4 was placed as more similar than II/12 because the harmony features one chord in each bar, alternating in a tonic-dominant pattern, using a very similar harmonic rhythm to the theme.

We can also measure harmonic similarity using a method of computational chord estimation, such as the methods utilised in the MIREX competition 'Audio Chord Estimation' (ACE) track.[30] For this study, the sonic-visualiser vamp-plugin 'NNLS Chroma' was used to compute the chromagrams from the audio files used in the study. NNLS Chroma analyses a single channel of audio and produces a tuning-adjusted chromagram. As the extracts have two channels, it calculates the mean of the two inputs. All the algorithm parameters used the 'generic popular song' suggestion, which emphasises the 'medium note range'.[31] I used this setting as there are no recommended settings for classical piano music.

30. Downie, "The Music Information Retrieval Evaluation Exchange (2005-2007): A Window into Music Information Retrieval Research" The Music Information Retrieval Evaluation eXchange (MIREX) is a community-based formal evaluation framework coordinated and managed by the International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL).

31. see http://www.isophonics.net/nnls-chroma for details on the different suggested settings.

FIGURE 2.23: The theme and four extracts (I/12, II/4, II/11, and II/12) from Brahms's Paganini Variations Op. 35 chosen for Question 4, with annotated Riemannian functions.

Similarity for this feature extraction method was determined using the chromagram of each extract and the theme. The chromagram visually represents the 'chromas' of a piece of music (a chroma is the pitch spelling or the pitch class, e.g. C, D, E, etc.). Each pitch class is associated with a numeric value, e.g. 0 refers to chroma C, 1 to C♯ and so on.[32] Using these numeric chroma's, we calculate the proportion of labels that are the same between the theme and the extract. I have chosen to use a standardised MIR evaluation software package — mir eval.[33] I specifically use their 'mirex' metric, as it compares the extracts based on each chord label sharing at least three pitch classes. Therefore, this method takes the two extracts' harmony from their chromagram's (one for each extract) and compares them using pitch-class methodology.

The visual chromagrams, such as for the theme (Figure 2.24), can be challenging to read, as they feature a variety of coloured blocks representing the intensity of a pitch. We might argue that for the theme there is a high intensity of colour for the pitches E, A and C, with other pitches occasionally featuring the same level of intensity. The pitches are represented on the vertical axis by the labels 'A♭', 'G' etc, going downwards. The horizontal axis represents the time through the extract, meaning the further we move away from the vertical axis the further through the extract we are. The coloured sections on the chromagram represent the intensity of that chroma (pitch-class) at different points in time. The key down the far left side (next to the pitch classes listed on the vertical axis) shows us that red represents the most intense (strongest presence) of the pitch, down to black which represents no presence of the pitch class; so for some parts, we have some pitches that are absent. The window increment used was 2048, with a window size of 16384 (as stated above, I used the standard settings).

To compute the harmonic similarity between the theme and an extract, we also need to create a chromagram of the extract we wish to compare with the theme; for example, extract I/12 (Figure 2.25). Again, visually, it could be described that the pitches E, A, and C are relatively prominent in this extract, but the pitch classes for A♭, F♯ and B also have prominence. To measure the similarity between the extracts, we take the output of this chromagram and the theme (a numerical table in CSV format that gives us the numbers for each section in time), and then map the theme and the extract using dynamic time warping (DTW).[34] This does not take into consideration a visual analysis of the chromagrams, but analyses the difference numerically between the chromas through time segments. This chapter uses the FastDTW algorithm with a Python implementation,[35] which finds the optimal alignment between two-time series through 'warping'

32. Meinard Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications* (Switzerland: Springer International Publishing, 2015).
33. mir eval is a python library that provides a transparent and easy-to-use implementation of the most common metrics used to measure the performance of MIR algorithms; see this web page for further detail https://github.com/craffel/mir_eval
34. see Chapter 3, Section 3.1.2, of Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications* for a detailed discussion of this process
35. See the following link for the python implementation: https://pypi.org/project/fastdtw/

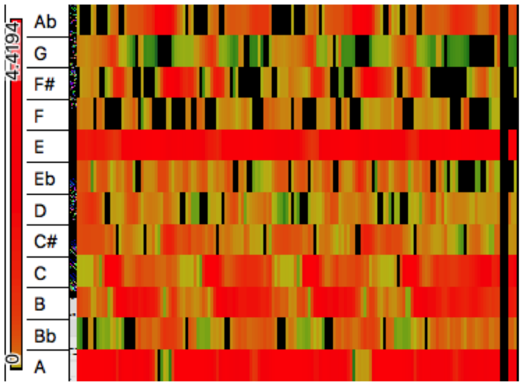


FIGURE 2.24: A chromagram of the first eight bars of the theme, from Brahms's Paganini Variations Op. 35. The vertical axis represents the pitch, the horizontal axis is time, and the intensity of the pitch at a set time in the audio is shown by the intensity of colour (red indicates the highest intensity).

the variation extract's time series, non-linearly, by stretching or shrinking the extract along its time axis until it matches the time series of the theme.[36] Once time-aligned, the cosine distance between each chroma in the theme and variation are calculated, using the MIREX weighted chord symbol recall metrics.[37] The 'mirex' vocabulary shows a 20% similarity level between the theme and variation I/12.

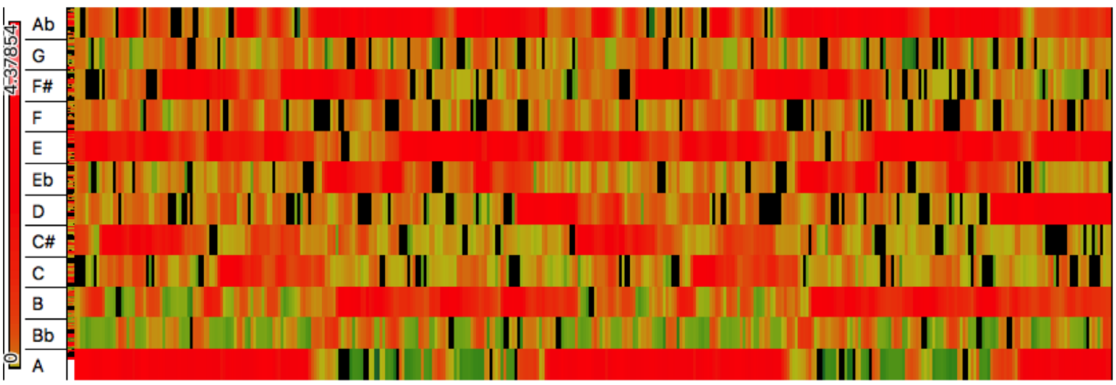


FIGURE 2.25: A chromagram of the first eight bars of Book I Variation 12, from Brahms's Paganini Variations Op. 35. The vertical axis represents the pitches, the horizontal axis is time, and the intensity of the pitch at a set time in the audio shown by the intensity of colour (red indicates the highest intensity).

36. Stan Salvador and Phillip Chan, "FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space," *Intelligent Data Analysis* 11, no. 5 (2007): 561–580.

37. More information on the MIREX weighted chord symbol recall metrics is available here: `http://craffel.github.io/mir_eval`

Overall, using MIREX weighted chord symbol recall metrics to determine a similarity level using the 'mirex' vocabulary (in the same method as above), I determine the following predicted rank ordering for Question 4: II/11 (28% similarity), I/12 (20% similarity), II/12 (19% similarity), and II/4 (10% similarity). Table 2.13 shows the rank positions for the two methods for each extract in this question.

Rank position	Music Theoretical	Non Theoretical
1	II/11	II/11
2	II/4	I/12
3	II/12	II/12
4	I/12	II/4

TABLE 2.13: The predicted rank order of similarity to the theme based on theoretical and non-theoretical approaches to harmony (Question 4), with 1 being the most similar to the theme, and 4 the least.

2.7.2 Results

The participants strongly agree that II/11 was the most similar extract for Question 4 (with 73% (117/162) agreement). The rest of the rank positions share a lower level of consensus (see Table 2.14 and Figure 2.26). The average rank positions for each extract show the ordering II/11 (1.44), I/12 (2.78), II/12 (2.86) and II/4 (2.91). There was no significant relationship between the predicted rank-ordering determined by Riemannian theory ($t(161)=24.81$, $p=0.05$) or by harmonic feature extraction ($t(161)=24.81$, $p=0.05$) and the population (for Riemannian theory $M=3.8$, $SD=2$, and feature extraction $M=3.51$, $SD=2.03$), using a one-tail t-test. However, Riemannian theory correctly predicted the extract that the population found most similar (rank position 1), and harmonic feature extraction correctly predicted the rank ordering most popular for all rank positions.

Table 2.15 shows the mean absolute differences for each ensemble category, as compared to the predicted rank orderings from Riemannian theory and feature extraction methods (Table 2.13). Figures 2.27 and 2.28 provide an observation of the proportion of participants that agreed with the extract predicted in each rank position. These graphs show no clear category that aligns most with either Riemannian theory or harmonic feature

Rank	I/12	II/11	II/12	II/4
1	10%	72%	8%	10%
2	35%	15%	26%	25%
3	23%	9%	38%	30%
4	33%	4%	28%	36%

TABLE 2.14: The proportion of participants that voted for each extract in each rank position, for Question 4. This table highlights the extract with the highest proportion of the participants for each rank position (1–4). The total number of participants is 162.

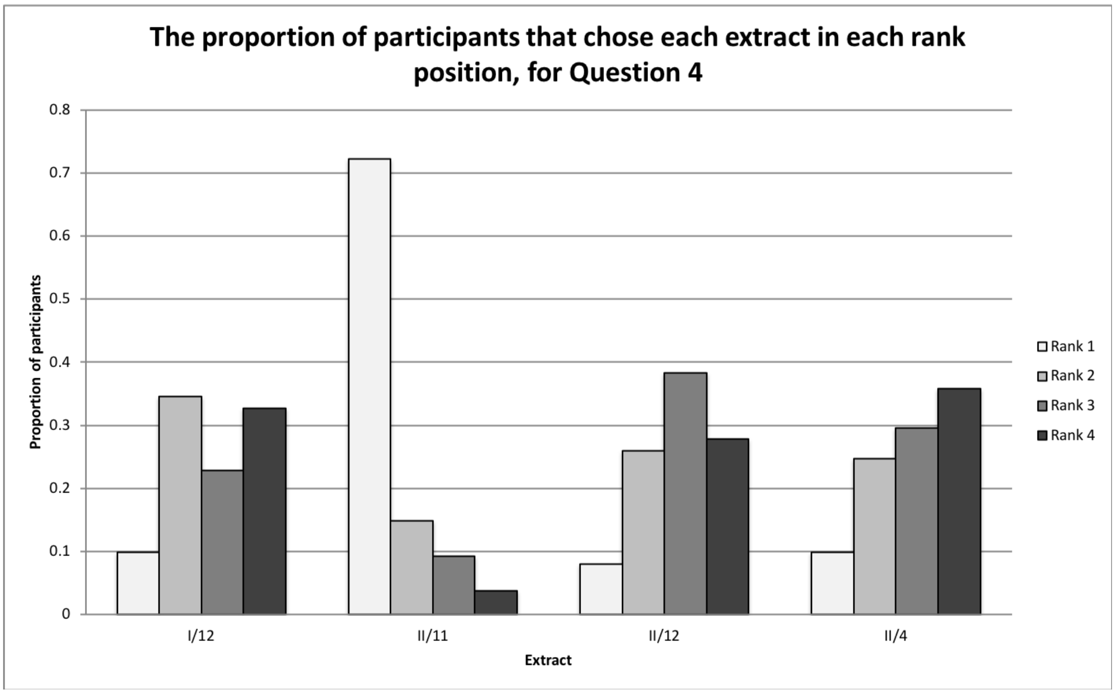


FIGURE 2.26: The proportion of participants that chose each extract in each rank position, for Question 4. The total number of participants was 162.

Demographic	**Riemann Mean Absolute Difference**	**Feature Mean Absolute Difference**
Has ensemble experience	3.5	3
Does not have ensemble experience	3.41	3.59

TABLE 2.15: The mean of the absolute differences between the score-based and audio-based approaches predicted rank ordering, for Question 4, grouped by ensemble experience. There were 103 participants with ensemble experience and 59 without experience.

extraction. Table 2.15, shows the mean absolute differences for the different ensemble categories: 'does not have ensemble experience' aligns most closely with Riemannian theory methods of determining similarity (M=3.41, SD=2.06); and 'has ensemble experience' aligns most closely with harmonic feature extraction methods of determining similarity (M=3, SD=1.79). The results of the t-test analysis in both scenarios revealed the difference between the means of having and not having ensemble experience was significant for only the feature extraction method (t(161)=-1.79, p=0.05). Thus, having ensemble experience made a participant significantly more likely to rank the extracts in a manner more aligned with harmonic feature extraction methods of determining similarity.

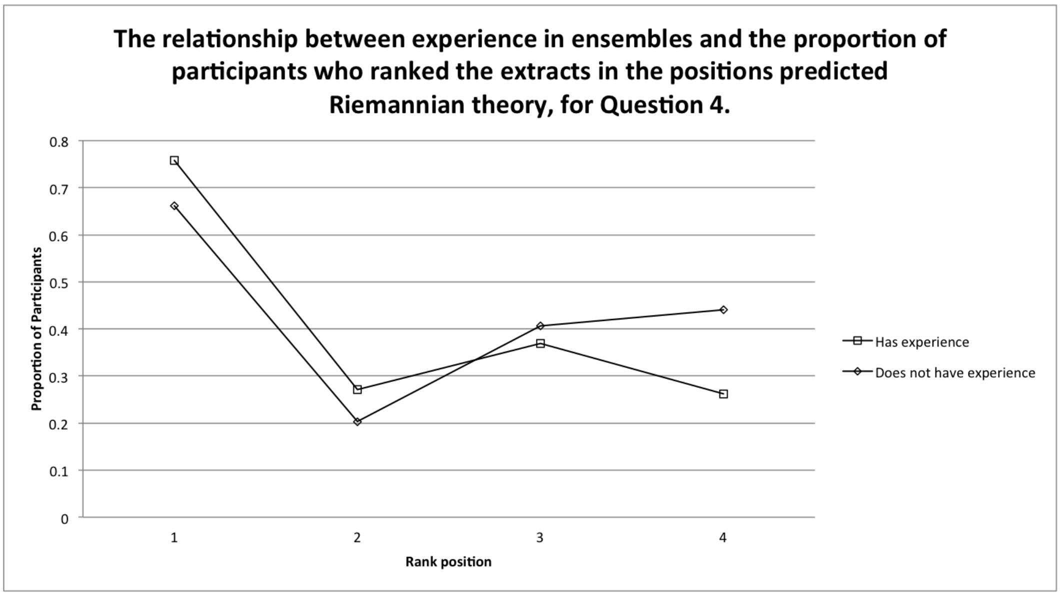


FIGURE 2.27: The proportion of participants that agreed with the rank order predicted by Riemannian theory for each extract in each rank position, for Question 4, grouped according to their ensemble experience. There were 103 participants with ensemble experience and 59 without experience.

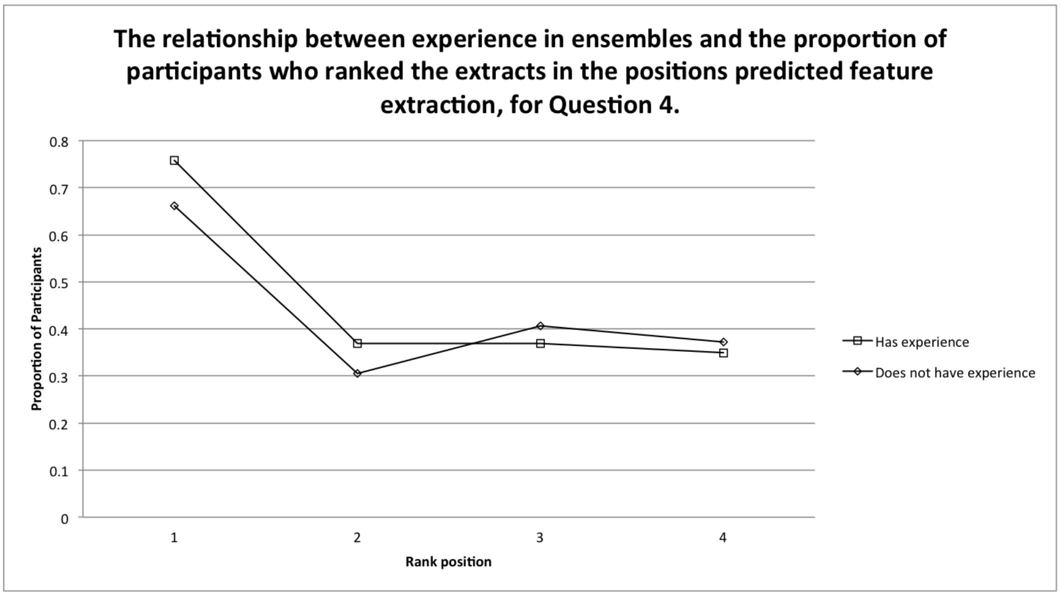


FIGURE 2.28: The proportion of participants that agreed with the rank order predicted by feature extraction techniques for each extract in each rank position, for Question 4, grouped according to their ensemble experience. There were 103 participants with ensemble experience and 59 without experience.

2.8 Question 5

2.8.1 Materials

Question 5 used no new material, and instead compared what each participant judged as most similar for the previous five questions. Each participant received the extracts they had placed in 'rank position 1' for Questions 1–4. I did not inform the participants that this is why they had these four extracts; the question told them to rank these extracts in the same manner as they had done for the previous four questions. There is no 'prediction' for this question; instead, I collected this data to determine what participants consider the most aurally important marker of similarity.

2.8.2 Results

This section will give an overview of the extracts prominently chosen in each rank position along with a comparison of the extracts as grouped by the question that the extracts were taken from (for example, extract II/9 was used initially in Question 1, which focused on form). This question showed that the theoretical approach *Formenhlre* aligned most with what participants determined as the most similar extract to the theme.

Overall, 29% of the participants picked extract II/9 as the most similar (rank position 1). Question 5 only presented the extracts that the participants ranked in position 1 for the first four questions. Therefore, II/9 would only have been available in Question 5 for those participants who had ranked it in position 1 in Question 1. Of those participants, 47% ranked it in position 1 (47/101). The second most voted for extract in rank position 1 was II/5, from the melody question (Question 2), with 22% of participants choosing it. Of those participants who had the extract to choose from, 43% ranked it in position 1.

No other rank position had similar levels of agreement across the whole population. II/11 was the most popular extract for the rank positions: 2 (16%), 3 (17%) and 4 (15%). Likely, this occurred because 117 participants would have had this extract to rank in Question 5, since in Question 4 of this study 73% of the participants (117 participants) agreed it was the most similar.

By grouping the extracts according to the question from the study they originate from (see Table 2.16), it was found that extracts from Question 2 (melody) were the most likely to be ranked in rank position 1 (36%). This question (Question 2) is a different question to the most popular extract in rank position 1 (II/9), which originated from Question 1 (form). However, as already noted, 22% of the whole population ranked extract II/5, from the melody question, as most similar — only 7% less than extract I/9 from Question 1. For mode rank positions for 2–4 there was not as strong agreement

Rank	Q1	Q2	Q3	Q4
1	35%	36%	10%	13%
2	17%	20%	22%	20%
3	16%	19%	24%	21%
4	18%	13%	26%	18%

Table 2.16: The proportion of participants that voted for an extract from each question in each rank position, in Question 5. This table highlights the extract with the highest proportion of the participants for each rank position (1–4). The total number of participants was 162.

on the question; the mode for these rank positions are all Question 3: 2 (22%), 3 (24%) and 4 (26%).

2.9 Discussion and Conclusions

This listening study observed whether music-theoretical methods could better predict which extracts a participant would find similar to a theme than feature extraction methods. I aimed to find out which (if any) traditional music analysis techniques could predict auditory perception of similarity. This study showed throughout that music-theoretical techniques could, at least, predict the extract that participants would find most similar to the theme. In contrast, feature extraction methods could not all predict the most similar extract (with the feature extraction method for form predicting none of the extracts' rank positions correctly).

Overall, the method with the highest level of alignment with participants' judgements of audible similarity was the melody-based feature extraction (with an absolute difference of 3.41), aligning 61% with the participants' rankings. Robert Welker's (1982) research has shown the importance of melody in theme and variation similarity judgements, finding that the theme's melody is the central tendency of variations.[38] In total, 20 participants (12%) agreed with the full rank order predicted by melodic analysis. Similarly, Riemannian theory also had 20 participants who agreed with the full rank order predicted by the theoretical approach. Harmony, both as the harmonic feature extraction method (aligning with 60% of participants' rankings, with an absolute mean difference of 3.51) and Riemannian theory (aligning with 57% of participants' rankings, with an absolute mean difference of 3.8), also aligned closely with participants' rankings. The most significant number of participants to agree on the full predicted rank ordering for a question was 27, for the harmonic feature extraction predicted rank ordering.

Form, as defined by Caplin's theory of formal functions, was the next most aligned method of determining similarity in participants' rankings — aligning 49%. Interestingly,

38. Welker, "Abstraction of Themes from Melodic Variations."

the approaches that did not seem to align closely with participants' similarity judgements were the Segmenter analysis of form (35%), and Schenkerian analysis (29%).

Stronger levels of agreement between the participants' rankings were found for the most and least similar extracts to the theme, across all the questions. This agreement was especially strong for Question 2, with I/9 in rank position 4 with 73% agreement, and Question 4 with II/11 in rank position 1 with 73% agreement. Other high levels of agreement included II/9 in rank position 1 for Question 1 (62%), II/1 in rank position 4 for Question 1 (55%), and II/5 in rank position 4 for Question 2 (52%). Across the first four questions, the demographic feature of having or not having ensemble experience produced a statistically significant difference in the means, for all but Schenkerian analysis. Most frequently, this showed that having ensemble experience made a participant significantly more likely to rank extracts in a manner that aligned more closely with the predicted rank ordering: namely for Question 1 (form) using the formal analysis and the Segmenter algorithm, Question 2 (melody) and Question 4 (Harmony) for only the feature extraction approach. For Question 4, when using Riemannian theory, the opposite was returned — not having ensemble experience made a participant significantly more likely to rank extracts in a manner that aligned with the predictions made by Riemannian theory. We could expect participants who have ensemble experience to be used to listening at a deeper level, and to hearing multiple streams or parts, which could explain why those with ensemble experience were more aligned with form and harmony-based approaches of determining similarity. However, the reverse may be equally valid: musicians who play in ensembles are more likely to be 'single-line' instrumentalists/singers (such as singing a vocal melody, or playing a violin line), whereas those who are not part of ensembles, yet still play musical instruments, may be used to thinking in multiple parts (such as pianists, classical guitarists etc.). This could explain why those with ensemble experience were more likely to align with melodic methods of determining similarity, and those without ensemble experience were more likely to align with Riemannian theory's definition of similarity. Therefore, participants who play different instruments may have different methods of listening, and the instrument that a musician plays could affect the applicability of different music-theoretical models to their decisions about similarity.

Question 5 measured if there was a prominent method of analysis that aligned with audible perceptions of similarity. This question did not have a consistent or reliable pattern. 47% of the participants (who had the extract to choose from) ranked extract II/9 in rank position 1. When grouping the extracts according to the question from the study that they originated from, Question 2 was the most popular question for rank position 1 (36%). Extract II/9 was originally from Question 1; therefore, either Question 1 or 2 could align most closely with the participants' perceptions of audible similarity. Due to issues with 'atomising' different aspects of music (form, melodic contour, voice leading, harmony), we cannot know whether there is a relationship between form (Question 1) or melody (Question 2) and auditory perceptions of similarity. The participants were unaware of the analytical method examined in each question, and, thus, we cannot assume that the participants made their judgements of similarity based on the analytical method proposed for each question. Instead, a participant could have used several different analytical or ad hoc methods for judging similarity. The extracts could not be repeated in different questions to avoid familiarity bias and thus did not enable me to compare the effect of the other extracts present in a question on the participants' ranking. We should, therefore, be wary of concluding, at this stage, that melody is the most crucial determinant for similarity, rather than, for example, harmony. In Question 1 form was observed, but the segments could be distinguished using either melodic or harmonic similarity, and in many instances in these extracts, the two go hand-in-hand. A more explicit indication of the reasons behind the participants' decisions, or further questions asking them to judge similarity based on a specific element, may be needed to address this shortcoming. A combination of all the music-analytical methods, making one single prediction of similarity, could be another method of overcoming this shortfall.

Upon reflection, musical expertise needs further detailing to enable a better exploration of this demographic. Ensemble experience could also be incorporated into this demographic feature to observe its effect on the expertise of a participant. Other information such as the amount of regular practice an individual participant undertakes; their participation in hobbies that use music, for example attending church services and singing every Sunday; current participation in ensembles; and what type of ensembles they are playing in, could all be important to this demographic features. It may also be relevant to distinguish between focused listening and 'background music'. Listening to music while working may have a different effect on a participant than the effect of the participant spending an hour a day on focused listening. For this study, the exposure a participant has had to classical music could also have affected their judgements.

In summary, similarity, as determined by melodic feature extraction, is the closest aligned to the auditory perception of similarity. The best performing music-theoretical approach (with respect to predicting participants' similarity judgements) appears to be Riemannian theory, which aligned with the same number of participants as melody. Question 4 (which considered Riemannian theory) saw the highest agreement on rank position 1, with extract II/1 ranked by 73% of participants in this position, in line with the Riemannian prediction. Additionally, the question that focused on harmony featured the highest number of participants to agree with the full rank order from the predictions. The results of this study, therefore, require a narrowing of the focus of this thesis in order to explore harmonic or melodic similarity. I have chosen to focus on harmonic similarity, as it has received less focus in scholarly research to date, whereas, melodic similarity has been explored extensively — both computationally,[39] and in the

39. for example, see Typke, Wiering, and Veltkamp, "Transportation Distances and Human Perception of Melodic Similarity"; Anja Volk and Peter Van Kranenburg, "Melodic Similarity Among Folk Songs: An Annotation Study on Similarity-Based Categorization in Music," *Musicae Scientiae* (London) 16, no. 3 (2012): 317–339; Berit Janssen, Peter van Kranenburg, and Anja Volk, "Finding occurrences of

field of music perception.[40] Furthermore almost no literature has observed how harmonic music-theoretical approaches (such as Riemannian theory) may aid in explaining musical similarity, and, analogously, Riemannian theory's computational effort has been limited (see Section 1.4.1 of Chapter 1). This study therefore highlights the possible alignment of determining harmony with auditory perception with music-theoretical approaches. I will explore this through examining datasets of aural harmonic transcription studies, and through my own aural harmonic transcription study. I hypothesise that if one annotator can perceive a different chord to another annotator, then the two chords could be seen as perceptually similar, as they are audibly mistakable. These studies will confirm that Riemannian theory can explain an element of harmonic similarity, and could provide a further explanation for developing chord extraction techniques. Furthermore, I will show that there is a perceptual similarity between chords related by a Riemannian harmonic function.

melodic segments in folk songs employing symbolic similarity measures," *Journal of New Music Research* 46, no. 2 (2017): 118–134

40. for example, see Zohar Eitan and Roni Y. Granot, "Primary versus Secondary Musical Parameters and the Classification of Melodic Motives," *Musicae Scientiae* Discussion Forum 48 (2009): 139–179; Eleanor Selfridge-Field, "Social Dimensions of Melodic Identity, Cognition and Association," *Musicae Scientiae* (London) Discussion Forum 4A (2007): 77–97; Sven Ahlbäck, "Melodic Similarity as a Determinant of Melody Structure," *Musicae Scientiae* Discussion Forum 4A (2007): 253–280; Müllensiefen and Frieler, "Modelling Experts' Notions of Melodic Similarity"; Welker, "Abstraction of Themes from Melodic Variations"

Part II

Chapter 3

Explaining harmonic inter-annotator disagreement using Hugo Riemann's theory of harmonic functions

3.1 Introduction

Music transcription by ear relies heavily on subjective perceptions of musical structures.[1] It relies upon an annotator's perception of what they hear, along with the process of extracting grammatically feasible harmonic constructs from those audio cues. The subjective nature of perception can lead to disagreements between annotators on what is the 'correct' transcription. Using the ear to transcribe harmony creates many subjective attributes, both concerning the determination of the component pitches, and in the separation of their overtone partials.[2] The vast number of heterogeneous transcriptions available through online repositories exemplifies these disagreements, such as in the Ultimate Guitar Repository.[3]

1. This chapter is an extension of the work discussed in my recent publication: Selway et al., "Explaining Harmonic Inter-Annotator Disagreement using Hugo Riemann's theory of 'Harmonic Function'" I would like to thank my co-authors of this paper for their collaborative help in the research and data-gathering used in this chapter. Nazir A. Jairazbhoy, "The 'Objective' and Subjective View in Music Transcription," *Ethnomusicology* 21, no. 2 (1977): 263–273; Anssi Klapuri, "Introduction to Music Transcription," in *Signal Processing Methods for Music Transcription*, ed. Anssi Klapuri and Manuel Davy (New York: Springer US, 2006), 3–20

2. Klapuri, "Introduction to Music Transcription."

3. https://www.ultimate-guitar.com/, Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music"

Inter-annotator agreement refers to the extent that human annotators concur. These metrics aim to measure the amount of homogeneity or consensus between different annotators. Previously, harmonic transcription tasks have featured only the consensus of annotators, aiming for high inter-annotator agreement. However, the need to better understand the nature of inter-annotator disagreement has led to the development of datasets containing multiple reference annotations, such as the Chordify Annotator Subjectivity Dataset (CASD) introduced by Vincent Koops et al. (2018),[4] the Rock Corpus introduced by Trevor De Clercq and David Temprely (2011),[5] and the dataset used in Yizhao Ni et al. (2013).[6] The research involving these datasets commonly aims to find an empirical upper bound for harmonic annotators' inter-annotator agreement. However, research exploring harmonic disagreement between annotators is in its infancy: in music-theoretical studies, this research has seen authors focus on comparing their analyses, such as the comparative analysis of the Rock Corpus by the two authors of Trevor De Clercq and David Temprely (2011), as a result of being limited to a minimal and bias dataset.[7] Meanwhile, MIR studies often examine the analyses of a relatively small number of songs, for example in the work of Yizhao Ni et al. (2013).[8]

The metrics used to measure annotator disagreement in MIR studies commonly focus on pitch-class agreement, i.e. the amount of pitch-class overlap among the annotators' chord labels for a particular segment. Arguably, this agreement occurs at the lowest level of abstraction, meaning we are purely observing whether the notes on the surface are the same. In contrast, music theories such as Hugo Riemann's theory of harmonic functions explain harmony on a more abstract level. Riemann's theory enables us to ascertain similarity through establishing which chords have the same harmonic function, i.e. which chords are 'substitutable'. Therefore, by adopting Riemannian theory, we can attend to the harmonic function of a chord, enabling us to establish links between chords intuitively perceived as similar in music (e.g. the relative major or minor).[9]

Following on from Chapter 2, this chapter explores how music theory, particularly Riemann's theory of harmonic functions, can help explain the apparent disagreements in human-annotated datasets of harmonic transcriptions, specifically CASD. This chapter aims to explore whether music theory can explain some harmonic inter-annotator disagreement. As a result of the findings in the previous chapter, which suggests that Riemannian theory can explain some of the annotator disagreement present in CASD, this chapter shows that a perceptual similarity is present between these chords. Furthermore,

4. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."

5. Trevor de Clercq and David Temperley, "A corpus analysis of rock harmony.," *Popular Music* 30, no. 1 (2011): 47–70.

6. Yizhao Ni et al., "Understanding Effects of Subjectivity in Measuring Chord Estimation Accuracy," *IEEE Transactions on Audio, Speech, and Language Processing* 21, no. 12 (2013): 2607–2615.

7. Clercq and Temperley, "A corpus analysis of rock harmony."

8. Ni et al., "Understanding Effects of Subjectivity in Measuring Chord Estimation Accuracy."

9. Krumhansl, Bharucha, and Kessler, "Perceived Harmonic Structure of Chords in Three Related Musical Keys"; Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations."

this chapter follows on from Chapter 2 by suggesting that Riemannian theory is related to audible similarity. This agreement is on a more abstract music-theoretical level, and can potentially show a higher level of agreement at this more musically informed harmonic function level. This research provides a new application of Riemannian theory, warranting the exploration of the theory's relationship to music perception. This chapter, therefore, suggests that if annotator disagreement is explainable using Riemannian theory, then Riemannian theory may reflect a form of perceived harmonic similarity.

3.1.1 Taking a Music Theoretical Approach

Transcribing harmony by ear relies on consciously acquired and specific musical domain knowledge. Harmonic analysis is often performed in relation to a symbolic representation, such as sheet music, meaning that an analysis based on the transcription of music by ear is often neglected. This skill relies heavily on personal subjective perceptions — firstly in terms of assessing the auditory cues, and secondly in their translation into musical structures — along with significant skill, increasing the propensity for subjective influence and leading to annotator disagreement.

Music-theoretical discourse on popular music regularly illustrates annotator disagreement (or inter-annotator disagreement), as recording practices often lead to a lack of notated music. Therefore, creating a transcription of a popular music song requires an annotator to perform a harmonic analysis by ear, and decide on the chord that best matches a particular segment. The literature surrounding the first chord of The Beatles song 'A Hard Day's Night' provides one such example. Ever since its recording, music theorists, experts and amateurs have tried to unravel the sound into its respective pitches, contributing to its 'holy grail' status as 'one of popular music's great unsolved mysteries'.[10] The complex cluster chord, with no 'original' notated version, has been perceived in a number of ways: as a G major chord, with suspended 4th, added 7th, 9th or 11th along with their inversions;[11] as an F major chord with added G and D;[12] and has even been associated with more complicated labels surrounding the function of the chord such as the dominant 9th of F,[13] the polytriad i7/5 in A♭ major,[14] a polychord which juxtaposes the tonic and subtonic.[15] The band members are also said to

10. Dominic Pedler, *The Songwriting Secrets of The Beatles* (London: Omnibus Press, 2003).

11. Tetsuya Fujita et al., *The Beatles: complete scores.* (Hal Leonard Publishing, 1993), 1136; Andrew Hickey, *The Beatles In Mono* (United States: Lulu Press, 2010); Bob Spitz, *The Beatles: the Biography* (London: Little, Brown / Company, 2005); Pedler, *The Songwriting Secrets of The Beatles*.

12. John C. Winn, *Way Beyond Compare: The Beatles' Recorded Legacy, Volume One, 1957-1965* (Crown Archetype, 2008); Kenneth Womack, *Maximum Volume: The Life of The Beatles Producer George Martin, The Early Years, 1926–1966* (Chicago Review Press, 2017).

13. Wilfrid Mellers, *Twilight of the Gods; the Music of The Beatles*, Richard Seaver Bks (New York: Viking Press, 1974), 215.

14. Steven Clark Porter, *Rhythm and Harmony in the Music of The Beatles* (City University of New York, 1983), 886.

15. Terence J. O'Grady, *The Beatles, a Musical Evolution*, Music Series (United States: Twayne Publishers Inc., 1983).

have disagreed on the chord label.[16] In interviews, The Beatles have given a mixture of partial or even conflicting views on what this chord is. For example, George Harrison perceived it as an F chord with a G on top, and then later when discussing it with Gary Moore, Harrison described the chord as a G7sus4.[17] This chord, therefore, embodies how harmonic transcriptions by ear can lead to annotator disagreement.[18]

In the west, music-theoretical approaches to harmony were developed for tonal art music, although, researchers have debated in favour of their appropriateness as a method for analysing popular (and vernacular) music.[19] For this chapter, I have chosen to utilise Hugo Riemann's theory of harmonic functions due to the results of my previous study (Chapter 2) which showed an alignment with audible perceptions of similarity and Riemann's construct of harmonic function. Riemannian theory's appropriateness for the analysis of popular music has been addressed by music scholars, especially for corpora of songs using traditional triad based harmony, thus featuring the tonic, subdominant and dominant chords or functions prominently.[20] Further research should utilise European art music (the genre that was utilised when creating the theory) to see if this also highlights a perceptual relationship between harmonic disagreement and Riemannian theory.

Riemannian theory also lends itself well to the study of similarity, though little work has explored how harmonic similarity can be related to function theory (see Eyton Agmon's work on revisiting harmonic function using prototype theory).[21] However, like many music theories, it was created with the concept of musical perception at its core.[22] Riemannian theory also lends itself well to this chapter's analysis as harmonising through substitutions is prominent in instrumental, improvisation, and composition teaching. Research has shown that trained musicians (such as improvisers) perceive musical structures with related functions as sounding similar.[23] As the annotators of this study are

16. Koops, "Computational Modelling of Variance in Musical Harmony."

17. Spitz, *The Beatles: the Biography*; Koops, "Computational Modelling of Variance in Musical Harmony."

18. For further discussions of the variants in the analysis of this chord, the reader is referred to Vincent Koops 2019.Koops, "Computational Modelling of Variance in Musical Harmony"

19. Nicole Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music," *Music Theory Spectrum* 32, no. 2 (2010): 95–110; Walter Everett, "Making Sense of Rock's Tonal Systems," *Music Theory Online* 10, no. 4 (2004); Christopher Doll, *Listening to Rock Harmony* (Columbia: Columbia University, 2007).

20. Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music"; Nicole Biamonte, "Modal Function in Rock and Heavy Metal Music," in *L'analyse musicale aujourd'hui*, ed. Modher. Ayari, Jean-Michel Bardex, and Xavier Hascher (Université: Delatour France, 2012), 275–290; Guy Capuzzo, "Neo-Riemannian Theory and the Analysis of Pop-Rock Music," *Music Theory Spectrum* 26, no. 2 (2004): 177–199; Doll, *Listening to Rock Harmony*.

21. Eytan Agmon, "Functional Harmony Revisited: A Prototype-Theoretic Approach," *Music Theory Spectrum* 17, no. 2 (1995): 196–214.

22. Riemann, *Harmony Simplified: or, the Theory of the Tonal Functions of Chords*; Riemann, "Ideas for a Study 'On the Imagination of Tone'"; Suzannah Clark, "On the Imagination of Tone in Schubert's Liedesend (D473), Trost (D523), and Gretchens Bitte (D564)," in *The Oxford Handbook of neo-Riemannian Music Theories*, ed. Edward Gollin and Alexander Rehding (New York: Oxford University Press, 2011), 294–321.

23. Andrew Goldman, Tyreek Jackson, and Paul Sajda, "Improvisation Experience Predicts how Musicians Categorize Musical Structures," *Psychology of Music*, 2018, 1–17.

trained musicians, this may go some way towards explaining annotator disagreement.

The musical score was required for each song in the dataset to assist this Riemannian analysis — used to compare the annotators' transcriptions.[24] It is important to note that most available scores for popular music are published notated arrangements (e.g. piano/vocal scores), and may themselves be subjective. Therefore, in this chapter, the scores provide cues for important motives, enabling the alignment of the lyrics with the harmony and providing the local and global keys. The score also enables the exploration of alternative causes of disagreement, for example, the level of granularity at which to annotate the harmony, and the presence of two chords simultaneously. Importantly, this study does not make a judgement on whether the harmony is correct or incorrect; instead, I aim to discern whether employing a music-theoretical approach can explain perceptual harmonic disagreements.

3.1.2 Dataset

This chapter uses the CASD dataset that was introduced by H. Vincent Koops et al. (2019) to study disagreement in harmony transcriptions.[25] This dataset contains chord labels from four different professional annotators of 50 songs from the McGill *Billboard* dataset.[26] Each *Billboard* annotation presents the harmonic annotation formed by a consensus of three or more experts in jazz and popular music. This dataset quickly became a standard reference set for several MIR tasks relating to harmony such as ACE (Automatic Chord Estimation). From this dataset, Koops et al. (2019) chose the 50 most played songs, based on the number of YouTube plays. At the time of collection, the least-played song in the dataset had 76,000 plays, and the most-played song had over 13 million.[27]

Koops et al. required their annotators to have completed formal study of music and harmony at undergraduate or graduate level, to have experience in performing (for example, in cover bands), and experience in transcribing popular music, to ensure high-quality transcriptions.[28] The four annotators chosen were successful professional musicians with a broad knowledge of harmony, who held academic degrees in music and had between 15 and 25 years of experience on their primary instrument — see Table 3.1 for an overview of the annotators. Half the annotators of this dataset were guitarists, and the other half were pianists; that is, they all play chordal instruments.

24. Arguably, a Riemannian analysis could be completed by ear if the analyst has an excellent ear and memory. However, this is not a skill held by the majority and, therefore, for most analysts (including myself) a score is required to complete a Riemannian analysis.

25. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."

26. Bradlow and Fader, "A Bayesian Lifetime Model for the 'Hot 100' Billboard songs"; Burgoyne, Wild, and Fujinaga, "An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis."

27. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."

28. Ibid.

Ann	**Instrument**	**Years**	**Background/ Occupation**	**Education**	**Annotation time (min)**	**Reported difficulty**	**Chord labels per song**
*A*1	Guitar	15	Transcriber, composer	Music Theory, Composition	23.10 (14.91)	2.40 (1.16)	9.46 (5.13)
*A*2	Guitar	19	Musician, teacher	Conservatoire	15.66 (9.91)	1.60 (1.18)	9.42 (4.20)
*A*3	Piano	25	Transcriber, composer	Piano, Composition	22.00 (7.42)	2.42 (0.73)	12.44 (5.83)
*A*4	Piano	20+	Composer, producer	Conservatoire, Composition	26.10 (12.18)	1.96 (1.07)	8.86 (4.70)

Table 3.1: Overview of annotators, their primary instrument, the number of years they have been playing this instrument, musical background, their musical education, average annotation time (in minutes), the annotators' reported difficulty in terms of how hard it was for them to annotate that song, and the number of chord labels per song. Difficulty is reported on a scale from 1 (easy) to 5 (hard). Standard Deviation is displayed in brackets. *Note:* table adapted from Hendrik Vincent Koops, "Computational Modelling of Variance in Musical Harmony" (PhD diss., Utrecht University, 2019).

To create CASD the annotators' work was a *task-focused* one: to listen to the music and transcribe the chord labels of the songs as they perceived them, so they could reproduce what they had heard. The annotators were provided with a web interface where they selected chord labels for each beat from a drop-down menu with all the chord labels that are available in *Billboard*. In the case that a chord label they wished to use was not available, the annotators notified the researchers and they added the chord label to the system. In this way, the annotators were free to choose any chord label for each beat, but it is worth acknowledging that this method could have introduced an undesirable delay for the annotators, meaning the vocabulary grew with the number of songs. Koops et al., reported that all of the annotators asked for chords to be added, suggesting that the annotators were not affected by this delay. The researchers maintained close contact with the annotators to enable them to request additions quickly.[29] The annotators were free to return to and edit their annotations throughout the study, meaning they could revise the annotations to make use of the extended vocabulary.

3.1.3 Disagreement

Koops et al. provide a detailed overview of the disagreement between annotators found in CASD.[30] Altogether, they found that each annotator used a particular set of chord labels — or vocabulary — for their transcriptions. Their vocabularies differed in size and content. That is, in addition to sharing standard chord labels in their transcriptions, each annotator used a subset of particular chord labels. These findings suggest the existence of a theoretical limit on human agreement in harmonic transcriptions. However, although each annotator used a particular chord-label vocabulary, the researchers found no statistically significant difference as to which annotator was most likely to disagree.[31] Interestingly, Koops et al. found that the chord label vocabularies used were more similar between the annotators who had the same primary instrument (e.g. between A_1 and A_2 who were both guitar players, and between A_3 and A_4 who were both pianists).[32] The results also suggested piano players were more diverse with their chord label vocabulary than the guitarists, though this cannot be generalised due to the small sample size.

Furthermore, in a pairwise analysis (i.e. a calculation of the average agreement between all possible pairs of annotators), Koops found that annotators disagreed on 24% of the chord base notes when in root position. This disagreement increased with the complexity of chord labels, to 41% when taking into account all pitch classes of the chords. This disagreement was even higher (an average 46%) when looking at inversions. In a comparable experiment using annotations from formally trained musicians, Yizhao Ni et

29. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."
30. Ibid.
31. Ibid.
32. Ibid.

al. (2013) reported approximately 10% disagreement among the annotators when compared to their consensus.[33] Similarly, De Clercq and Temperly (2011) reported a 7.6% disagreement rate.[34] In this study, the disagreement rate is much higher any 'errors' made by the annotators were not corrected. De Clercq and Temperly (2011) and Ni et al. (2013) claimed to correct errors that were unintentional before calculating their disagreement rate.[35] We (i.e. my co-authors and I) decided not to remove these 'errors', as this paper did not aim to infer a right or wrong harmony but rather to compare the disagreements between annotators, intentional or otherwise.

This chapter exploys a *global agreement* analysis, instead of the pairwise agreement performed by Koops et al.,[36] because our methodology aims to observe overall disagreement between annotators. Global agreement refers to assessing whether all annotators agree or disagree on the chord for agreement. This assessment highlights when all four annotators agree or disagree on the makeup of the chord. For this chapter, embellishments such as sustained chords, 7ths etc. are ignored. Therefore, for example, no distinction is made between C major and C major7. We decided to ignore embellishments because, using Riemannian theory (discussed in Chapter 1, Section 1.3.1 and expanded on in Section 3.2.1), these two chords would be perceived as having the same harmonic function, because they share the same root, 'C'. This is an appropriate methodology for the music in this dataset; however, if the dataset had included musical genres such as jazz, for which the use of embellishments is highly essential, this method of considering the underlying chord alone could be too simplistic. Therefore, this analysis was performed at the major/minor level, meaning, for example, that we acknowledged a difference between C major and C minor. At this level, we found a 34% global disagreement. This chapter will focus on this 34% disagreement, applying a music-theoretical approach in an attempt to explain it, and this will be discussed in the following section (3.2). This chapter provides an insight into whether Riemannian theory could explain harmonic disagreement.

3.2 Method

By aligning the scores with each annotator's audio-based annotations, we performed both a global analysis of the annotators' disagreements in CASD, and explored possible musical explanations for any remaining annotator disagreement. The scores were sourced mostly from *Musicnotes*,[37] due to its large catalogue (over 300,000 pieces), and its wide

33. Ni et al., "Understanding Effects of Subjectivity in Measuring Chord Estimation Accuracy."
34. Clercq and Temperley, "A corpus analysis of rock harmony."
35. Clercq and Temperley, "A corpus analysis of rock harmony."; Ni et al., "An end-to-end Machine Learning System for Harmonic Analysis of Music."
36. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."
37. Musicnotes: `https://www.musicnotes.com/`. Using sites Sheetmusicnow (`https://www.sheetmusicnow.com/`) and Sheetmusicplus (`https://www.sheetmusicplus.com/`) when the scores were not available through Musicnotes.

popularity as a source of sheet music. Sheet music was sourced for 41 of the 50 songs in the dataset; thus I reduce the dataset to these 41 songs. The nine scores not available were mostly covers, mash-ups, or from a musical practice that features improvisation, and were therefore unlikely to be notated. In Chapter 5, I will return to these nine songs and explore an alternative methodology, which removes the functional element of the proposed methodology (identifying function being the aspect which requires the score), and focus on the relationship between the chords using just substitutions.

It is important to note that the scores found for the songs in CASD are often published notated arrangements (e.g. piano/vocal scores), which may not have been created by the writers of the songs but by professional arrangers.[38] Thus, it could be that these are only an approximation of the song, or to paraphrase the words of Nicholas Cook: a symbolisation of the musical sound rather than a representation, as it may not match the rhythm and notes exactly.[39] These transcriptions themselves, therefore, may also suffer from the subjectivity of the transcriber. However, the scores do enable the analyst to see the prominent and distinctive musical features of the song, such as the main guitar riffs, vocal line and important harmonic features.

To allow for a comparison between the scores with the annotators' audio-based annotations, we aligned the per-beat chord labels with the specific beats of the bars in the score. The Chordify interface and beat tracker was initially used to create the original audio-based, per-beat chord label annotations of CASD. To improve the annotations, we used human beat tracking data to correct the Chordify beat tracking data manually. We obtained beat annotations by asking a different annotator to tap the beats of the song while it was playing. It is worth noting that the beat annotations could also be considered subjective.[40] To align the CASD annotations with the corrected beat annotations, we found for each score the closest matching beat in terms of real-time in the CASD chord label annotations. After repeating this process for each beat and for each annotator, we obtained beat-corrected chord label annotations for each of the annotators in CASD. Now with beat corrected chord labels, the beats were lined with each beat-per bar in the score — providing each beat with a bar number along with the beat (e.g. Bar 1 beat 1, Bar 1 beat 2 etc.).

For the remainder of Section 3.2, the methodology will be separated into two approaches. Firstly, I will discuss how Hugo Riemann's theory of harmonic functions will be utilised to explore annotator disagreement. Secondly, I will discuss other ways we utilised the musical score to explore the remaining harmonic disagreement between annotators, such as prolongation and harmonic ambiguity.

38. Clercq and Temperley, "A corpus analysis of rock harmony."
39. Nick Cook, "Towards the Compleat Musicologist?" (Keynote address at the 6th International Society for Music Information Retrieval (ISMIR) Conference, London, 2005).
40. Matthew Davies and Sebastian Böck, "Evaluating the Evaluation Measures for Beat Tracking," In *Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference* (Taipei), 2014, 637–642.

3.2.1 Riemannian Analysis

As discussed in Chapter 1 Section 1.3.1, one of the most prominent and influential theories of tonal harmony comes from the music-theoretical discourses of Hugo Riemann.[41] His theories move away from the traditional harmonic emphasis of a triad's relationship to the tonic, and instead focus on the harmonic purpose of a chord. Riemann states how many chords can assert the same harmonic function through acting as substitutions.[42] This means that similarity can be ascertained by identifying the chords with the same harmonic function i.e. the chords that are 'substitutable'. Thus, these harmonic functions establish links between chords previously assumed to be disparate and 'unsimilar', showing how two chords can have the same sense of function regardless of different pitch-class content.[43] This harmonic similarity implies that, on a functional level, some assumed annotator disagreement has a perceptual similarity present between the chords.

This study uses Riemann's three basic substitutions: the *Parallele*, *Variante*, and *Leittonwechsel*. In my methodology, I additionally utilise more complex substitutions, including the typical pop chord substitutions: the dominant of the dominant (`V` of `V`), and the subdominant of the subdominant (the 'backdoor cadence' — `IV` of `IV`)[44]. These substitutions are explained in Riemann's book on harmony,[45] which states that doubling the `S` or `D` function (`IV` of `IV` or `V` of `V`) can be understood as altered forms of the dominant and subdominant harmonies, turning them into their opposite — so `S` of `S` becomes `D`, and `D` of `D` becomes `S`. In F major, for example, the `S` of `S` is E♭ major, and `D` of `D` is G major,[46] see Figure 3.1. Thus in F major, E♭ can have a dominant function, and G major a subdominant function. Another complex substitution used for this study, is the understanding that diminished seventh chords can hold both a subdominant and dominant function.[47] The diminished seventh chord can replace the dominant seventh chord in a key (omitting the root): for example in C major, the diminished seventh chord of B–D–F–A♭ is similar to the dominant seventh chord G–B–D–F with the root omitted. Similarly, the diminished seventh also can take a subdominant function as the chord can either be seen as a sharpened `iv` (F–A♭–D) or we can see the pitches of the chord are prominently related to `Sp` with a flattened 5th (D–F–A♭).

41. Riemann, *Harmony Simplified: or, the Theory of the Tonal Functions of Chords*; Riemann, "Ideas for a Study 'On the Imagination of Tone'."
42. Harrison, *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of its Precedents*.
43. Ibid.
44. As described by Jazz theorist Coker (*Complete Method for Improvisation: For All Instruments*), the name derives from taking a different root to `I` than the usual `ii-V7-I`, replacing `V` with the progression `IV-bVII`
45. Riemann, *Harmony Simplified: or, the Theory of the Tonal Functions of Chords*.
46. Justin Hoffman, *Listening with Two Ears: Conflicting Perceptions of Space in Tonal Music* (Columbia: Columbia University, 2011).
47. Riemann, *Harmony Simplified: or, the Theory of the Tonal Functions of Chords*.

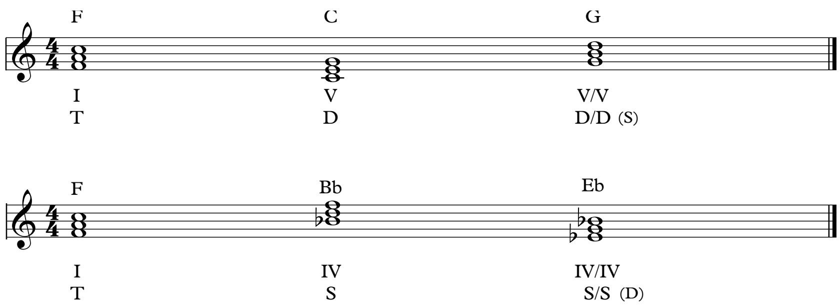


Figure 3.1: The subdominant of the subdominant (`S/S`) and dominant of the dominant (`D/D`) substitutions in F major, which demonstrate that `S/S` has a `D` function and `D/D` has a `S` function, as detailed by Hugo Riemann's theory of harmonic functions

Once function labels have been associated with each annotator's chord label, all annotators' functions for a beat are compared. Those beats where there is some annotator disagreement (the 34% detailed in Section 3.1.3) are then categorised as either `Agreement`, `Partial Agreement` or `No Agreement`. `Agreement` refers to the chord-label disagreements on which there is a full agreement on the Riemannian function. For example, assuming the key of C, the chord labels C, Am, Em, and C, would be analysed as `T`, `Tp`, `Tl`, `T` (i.e. Tonic, Tonic *Parallele*, Tonic *Leittonswechsel*, Tonic). Although they differ in their precise chord identity (`T`, `Tp`, `Tl`), they are all of tonic function. The chord labels are different, but the chords themselves are similar due to all of them being substitutions of the same function.

`Partial agreement` refers to a majority agreement in the function between the unique substitutions. I have utilised this category to account for subjectivity in harmonic annotation, allowing the exploration of situations of majority agreement. For example, in C major, the chord labels Am, A, Cm, G are analysed as `Tp`, `TP`, `t`, `D` (i.e. Tonic *Parallele*, the major *variante* of the Tonic *Parallele*, Tonic minor, Dominant). There is a majority agreement on the tonic function of the chord labels; three out of four chord labels are of a tonic function (`T` or `t`), while the outlier chord label G major has a dominant function.

`No Agreement` refers to the chord-label disagreements that have conflicting Riemannian functions, for example in the key of C major, the chord labels C, C, G7, and G are analysed as `T`, `T`, `D`, and `D` (Tonic, Tonic, Dominant and Dominant). Between the annotators, there is no majority agreement on function, as two annotators assigned the chord to a tonic function, and two to a dominant.

Although this may appear counter-intuitive, the categories `Partial Agreement` and `No Agreement` look only at the unique substitutions, and do not consider when one chord is dominant between the annotators. For example, `T`, `T`, `T`, `D` would be categorised as a `No Agreement`, because the unique substitutions `T` and `D` are not of the same function. However, there is a 75% agreement between the four annotators (three out of four agree on the function). This approach was chosen for this chapter as I am interested in whether it could explain the disagreements on chord labels between annotators. Current metrics already enable us to observe the similarity where there is a majority agreement on the same chord, such as the example just discussed. Future research could look at combining both methodologies to enable the exploration of all forms of annotator agreement/disagreement.

3.2.2 Score-based analysis

After completing the first stage of this methodology (identifying possible substitutions using Riemann's theory of harmonic functions), the aligned scores were examined to see if they could explain any of the remaining annotator disagreement. Firstly, I observed disagreements in the harmony that were caused by different instruments playing different chords. This method of explaining disagreement arose from the annotator instructions given by Koops et al. The task asked the annotators to transcribe the harmony of the song in a way that, in their view, best matched their instrument. Two annotators were pianists and two guitarists (see Table 3.1), and therefore the specific task, in combination with the annotator's primary instrument, was hypothesised as a potential cause of some harmonic disagreement.

Secondly, the scores allowed us to observe whether any remaining disagreement could be explained on the level of granularity, using Heinrich Schenker's concept of prolongation (please refer to Chapter 1 Section 1.3.4 for an explanation of Schenkerian analysis). This term refers to the elaboration, or 'composing out' of music's underlying structures.[48] In music theory, prolongation refers to a note that governs a span of music without necessarily sounding.[49] In Schenkerian analysis, we can, therefore, see a more complex structure made up of passing notes, arpeggios and other embellishments as being a simple prolongation of a single or a few notes at a different hierarchical level (see Figure 3.2 for an example of where a prolongation can affect the harmonic annotation of a passage). As these prolongations result from a transformation that turns notes at one level of Schenker's hierarchy into notes on another, they can be seen to create similarity by preserving sameness at one level, and introducing differences at others.[50] For example, in Figure 3.2, observing the harmony at a per-beat level, we see it change from C major to G major and return to C major. Instead, observing the harmony at a higher level, we can perceive the harmony of the whole bar as being in C major. In this chapter, I do not perform full Schenkerian reductions to reduce the pieces down to their Ursätze. Instead, the prominence of prolongations is observed within the annotators'

48. Drabkin, "Prolongation"; Cadwallader and Gagne, *Analysis of Tonal Music: A Schenkerian Approach*.
49. Forte and Gilbert, *Introduction to Schenkerian Analysis*; Pearsall, "Harmonic Progressions and Prolongation in Post-Tonal Music."; Drabkin, "Prolongation."
50. Larson, "The Problem of Prolongation in 'Tonal' Music: Terminology, Perception, and Expressive Meaning."; Forte and Gilbert, *Introduction to Schenkerian Analysis*.

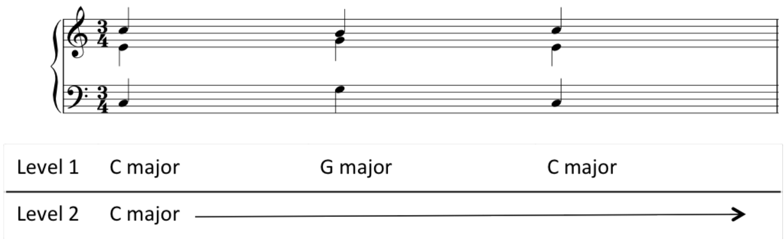


FIGURE 3.2: How the prolongation of notes over a bar can affect harmonic annotation depending on the hierarchical level at which the annotator is observing their annotation.

harmonic disagreement to see if this disagreement occurred due to differing perceptions of granularity.

3.3 Results: Example Analyses

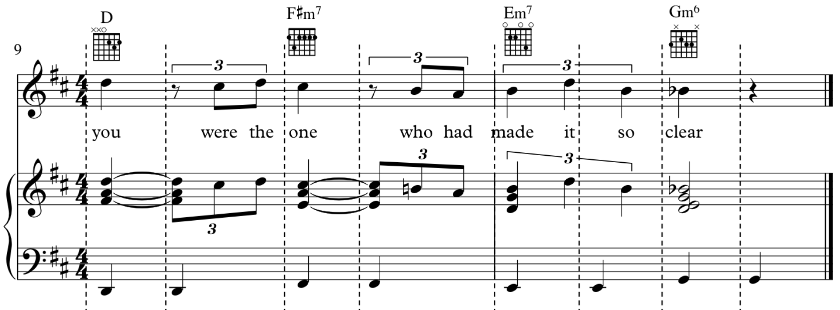
This section will provide five examples from a variety of songs in this dataset, showing a mixture of the different disagreement categories: `Agreement`, `Partial Agreement`, and `No Agreement` (defined in Section 3.2.1). The harmonic disagreements of each extract are presented, not only using Riemann's theory of harmonic functions, but also other methods of score-based analyses (as discussed in Section 3.2.2). Each example is, first, analysed in terms of Riemannian theory's ability to improve our understanding of disagreement, and second, any remaining features in the score that could explain the harmonic disagreement.[51]

3.3.1 'All Those Years Ago' by George Harrison

The first example is an extract from 'All Those Years Ago' by George Harrison (bars 9–10). For this song 77% of the disagreements share the same Riemannian function (`Agreement`), and a further 14% partly share the same function (`Partial Agreement`). Thus, in total, 91% of the disagreements in this song can be explained, at least partially. Figure 3.3 shows an example of where we can fully explain the annotators' disagreements using Riemann's theory (`Agreement`). The second half of bar 9 (the third and fourth beats of the diagram) provides an example of a disagreement that shares the same harmonic function (`Agreement`): A_1, A_3 and A_4 agree on the chord F♯ minor7 (`Tl`) for these two segments, whereas A_2 disagrees and believes that it is D major with the F♯ in the bass (`T`).

Interestingly, A_2 perceives a single chord in bar 9, making us consider whether A_2 is transcribing the harmony at a different level of granularity to the other annotators.

51. The examples chosen for this section happen to show A_4 disagreeing most frequently with the other annotators. As mentioned in section 3.1.3, A_4 was no more likely to disagree with the other annotators.



A_1	D:maj (T)	D:maj (T)	F#:min7 (Tl)	F#:min7 (Tl)	E:min7 (Sp)	E:min7 (Sp)	G:min6 (s)	G:min6 (s)
A_2	D:maj (T)	D:maj (T)	D:maj/3 (T)	D:maj/3 (T)	E:min (Sp)	E:min (Sp)	G:dim (S or sp)	G:dim (S or sp)
A_3	D:maj (T)	D:maj (T)	F#:min7 (Tl)	F#:min7 (Tl)	E:min (Sp)	E:min (Sp)	E:dim/b3 (dp or S)	E:dim/b3 (dP or S)
A_4	D:maj (T)	D:maj (T)	F#:min7 (Tl)	F#:min7 (Tl)	E:min7 (Sp)	E:min7 (Sp)	G:min6 (s)	G:min6 (s)

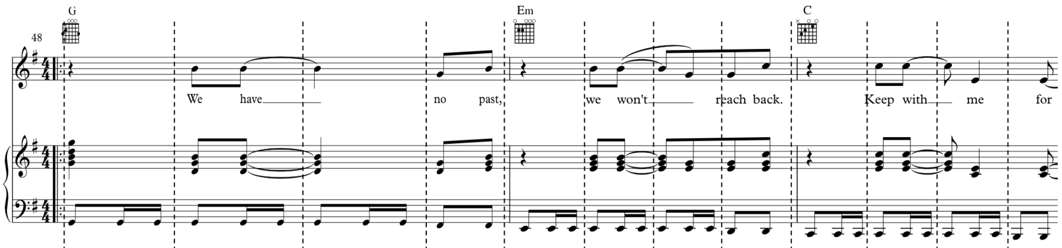
FIGURE 3.3: Bars 9 to 10 of 'All Those Years Ago' by George Harrison. This figure shows the musical score aligned with each annotator's per-beat chord label. We are interested in the use of the same harmonic function at the end of bar 9 and end of bar 10.

However, observing bar 10, A_2 does not continue their pattern of annotating a single chord per bar; therefore, this seems unlikely. Nonetheless, A_2 did not perceive the harmonic change that the other annotators highlighted in the second half of bar 9; only the introduction of a new root note. When looking at these chords without a function interpretation, the similarity between them is still apparent: the transcribers all agree the bass note is F♯, and the two chords share two common tones (out of four). Through using Riemannian theory, we can explain how F♯minor7 is assuming a tonic function as the tonic *Leittonwechsel* (`Tl`) of D major. Thus, in this context, either chord would be capable of performing the same harmonic function.

Later in this example, there is an agreement on the subdominant function, through more distantly related substitutions using diminished chords (second half of bar 10 in Figure 3.3). As explained in Section 3.2.1, diminished chords provide a dual function (often subdominant and dominant functions). A_2 perceives this chord as G diminished, and A_3 as E diminished/♭3. In D major, G diminished provides a subdominant function: both an `S` (due to the notes G and B♭ relating to E minor) and an `sp` substitution (B♭ and D♭ relating to B♭ minor). In contrast, E diminished/♭3 has dual function, as `dP` (pitches E and G relating to C major) and `S` (G and B♭ relating to E minor). Thus, all four annotators' chords have a subdominant function for their chosen harmony.

3.3.2 'All Through the Night' by Cyndi Lauper

Cyndi Lauper's song 'All Through the Night' demonstrates another clear example of the `Agreement` category. For this song, 68% of the disagreements can be explained



A_1	G:maj (T)	G:maj (T)	G:maj (T)	G:maj (T)	E:min (Tp)	E:min (Tp)	E:min (Tp)	E:min (Tp)	C:maj (S)	C:maj (S)	C:maj (S)	C:maj (S)
A_2	G:maj (T)	G:maj (T)	G:maj (T)	G:maj (T)	E:min (Tp)	E:min (Tp)	E:min (Tp)	E:min (Tp)	C:maj (S)	C:maj (S)	C:maj (S)	C:maj (S)
A_3	G:maj (T)	G:maj (T)	G:maj (T)	G:maj (T)	E:min7 (Tp)	E:min7 (Tp)	E:min7 (Tp)	E:min7 (Tp)	C:maj9 (S)	C:maj9 (S)	C:maj9 (S)	C:maj9 (S)
A_4	D:maj (D)	G:maj (T)	G:maj (T)	E:min (Tp)	E:min (Tp)	G:maj (T)	G:maj (T)	G:maj (T)	C:maj (S)	C:maj (S)	A:min (Sp)	A:min (Sp)
Disagreement category	N			A	A	A	A	A			A	A

FIGURE 3.4: Bars 48 to 50 of 'All Through the Night' by Cyndi Lauper. This figure shows the musical score aligned with each annotator's per-beat chord label. We are interested in the use of the same harmonic function at the end of bar 48, beats 2 to 4 of bar 49 and the second half of bar 50.

using Riemannian functions (`Agreement`), and a further 6% through partial matching of function (`Partial Agreement`). Figure 3.4, bars 48–50 of the score, shows three instances of agreement on a functional level. The first can be observed at the end of bar 48, where A_1, A_2 and A_3 agree on the chord G major (`T`), but A_4 perceives the chord as E minor (`Tp`). These two chords both perform a tonic function through the *Parallele* substitution. The second arises from beat two of bar 49, which has the same chords as the previous example, though more annotators (1–3) perceive E minor (`Tp`) and only A_4 disagrees and hears it as G major (`T`).

Finally, the same substitution performed within the subdominant function can be found in the second half of bar 50 (Figure 3.4), where A_1, A_2 and A_3 perceive the harmony as C major (`S`) and A_4 perceives it as A minor (`Sp`). This extract, of 'All Through the Night' also shows an example of where we cannot explain the annotator disagreement using Riemannian theory: bar 48 (`No Agreement`). The first beat of bar 48 shows harmonic disagreement over the function, with A_1, A_2 and A_3 perceiving the beat in G major (`T`) and A_4 perceiving it in D major (`D`). This disagreement, however, can be explained by the previous bar: bar 47 (not shown) finishes with a D major chord, and therefore it may be that A_4 still hears the harmony from the previous bar, or that the segments, as broken up by the beat annotator, overlap these two bars. This prolongation of harmony leads to disagreement over exactly where the harmony changes; importantly, there is agreement on the harmonic progression (`D`–`T`).

3.3.3 'Super Freak' by Rick James

Riemannian theory explains only 8% of the harmonic disagreements in Rick James' song 'Super Freak' as `Agreement` and a further 4% in the category of `Partial Agreement`.

Riemannian theory cannot explain the harmonic disagreement (`No Agreement`) in this example (bars 1 and 3 of Figure 3.5), but the disagreement is explainable by other musical features. The annotators disagree on whether the first two beats of each bar are in D major (`S`, with A_1 specifying the power chord D:5[52]) or A minor (`t`) as perceived by A_3. Thus, the annotators disagree on the function of the chord, between the subdominant and tonic functions.

The bass guitar riff present at the beginning of bars 1 and 3 is repeated nearly continuously throughout the song, and this is a prominent feature of the piece. However, we can observe a few explanations for this disagreement by examining the score. Firstly, the bass guitar part, as notated in the bottom stave (Figure 3.5), falls from the pitch D to an A, resembling a D major or minor chord (specifically a D:5). Then, the piano enters with an A minor chord, which is agreed upon by all annotators. Therefore, annotator A_3 is perceiving the harmony at a less granular level and taking the A minor harmony from beat 3 for the whole bar.

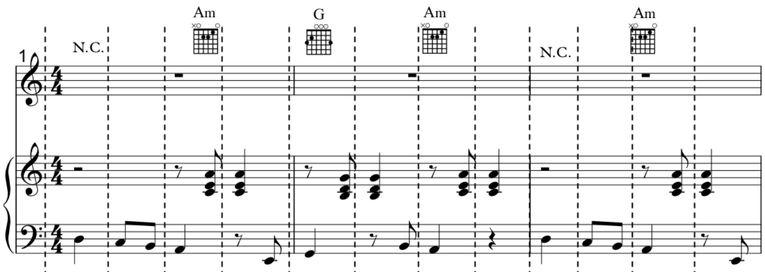
In contrast, the remaining annotators changed the harmony based on the arpeggio figure of the bass line. Interestingly, the guitar players in the dataset (A_1 and A_2) more often chose the chords that related to the guitar riffs in the piece, whereas the pianists chose the chords relating to the piano part or vocal line (for example, A_3 chose to follow the piano line, ignoring the bass line).

'N.C.', or 'no chord', is written above the stave over the guitar riff. This suggests, strictly speaking, that there is no harmony — it is a monophonic line. Therefore, the annotator who viewed the harmony as A minor, continuing the harmony of the proceeding and following beats, followed what this score implies. The conflict over whether to infer harmony during the bars annotated as 'N.C' causes 28% of this song's annotator disagreement. In total, both the score and Riemannian approaches explain 40% of the disagreements in this song.

3.3.4 'All Those Years Ago' by George Harrison, revisited

Returning to the piece 'All Those Years Ago' by George Harrison (the same piece used in Figure 3.3) provides another example of `No Agreement` (see bars 45 to 46 in Figure 3.6). A_4 disagrees with the other three annotators on the chord label for the last beat of bar 45: A_4 observed the chord D major5 (`T`), whereas the other annotators perceive it to still be in E minor (`Sp`). As the functions are different, we cannot explain this using Riemannian theory. The same is true in bar 46, where A_4 perceives the chord label to be D major/5 (`T`) and A_1, A_2 and A_3 perceives it as A major (`D`). Observing the score in more detail highlights how some annotator disagreement could have arisen

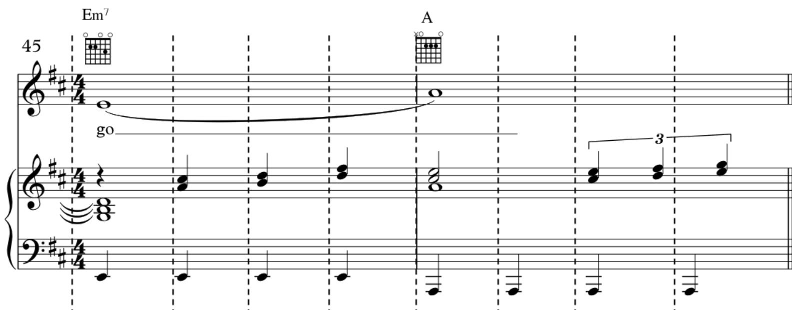
52. A power chord is a chord made up of only the first and fifth notes of the chord, removing the third, thus giving it neither a particularly major nor minor quality. The power chord is a widespread technique in popular music as the chord positioning allows for smooth transitioning between multiple keys.



A_1	D:5 (S)	D:5 (S)	A:min (t)	A:min (t)	G:maj (dP)	G:maj (dP)	A:min (t)	A:min (t)	D:5 (S)	D:5 (S)	A:min (t)	A:min (t)
A_2	D:maj (S)	D:maj (S)	A:min (t)	A:min (t)	G:maj (dP)	G:maj (dP)	A:min (t)	A:min (t)	D:maj (S)	D:maj (S)	A:min (t)	A:min (t)
A_3	A:min (t)	A:min (t)	A:min (t)	A:min (t)	G:maj (dP)	G:maj (dP)	A:min (t)	A:min (t)	A:min (t)	A:min (t)	A:min (t)	A:min (t)
A_4	D:maj (S)	D:maj (S)	A:min (t)	A:min (t)	G:maj (dP)	G:maj (dP)	A:min (t)	A:min (t)	D:maj (S)	D:maj (S)	A:min (t)	A:min (t)
Disagreement category	N	N							N	N		

FIGURE 3.5: Bars 1 to 3 of 'Super Freak' by Rick James. This figure shows the musical score aligned with each annotator's per-beat chord label. We are interested in the disagreement of the function of the chord at the beginning of bars 1 and 3.

through the annotators perceiving the harmony at different levels of granularity. For the fourth beat of bar 45, the annotators disagree between E minor and D major (with the different functions of S and T). In the score, the middle stave has a rising third pattern, which raises to a D and F♯ against the held E in the vocal and bass lines. Therefore, the disagreement appears to reflect the concept of granularity adopted from Schenker's concept of prolongation. In this example, A_4 has adopted a more granular approach, observing changes in harmony with any instrument's melodic movement. However, A_1, A_2 and A_3 took a broader view (less granular) of the harmony prolonging the chord with the held vocal and bass line, not changing the harmony with the melodic changes present in the inner voices. By observing the effects of granularity and Schenker's concept of prolongation, we can explain a further 2% of the annotators' disagreements.



A_1	E:min (Sp)	E:min (Sp)	E:min (Sp)	E:min (Sp)	A:maj (D)	A:maj (D)	A:maj (D)	A:maj (D)
A_2	E:min (Sp)	E:min (Sp)	E:min (Sp)	E:min (Sp)	A:maj (D)	A:maj (D)	A:maj (D)	A:maj (D)
A_3	E:min (Sp)	E:min (Sp)	E:min (Sp)	E:min (Sp)	A:maj (D)	A:maj (D)	A:maj (D)	A:maj (D)
A_4	E:min7/4 (Sp)	E:min7/4 (Sp)	E:min7/4 (Sp)	D:maj/5 (T)	D:maj/5 (T)	D:maj/5 (T)	D:maj/5 (T)	D:maj/5 (T)
Disagreement Category				N	N	N	N	N

FIGURE 3.6: Bars 45 and 46 of 'All Those Years Ago' by George Harrison. The figure shows the musical score aligned with each annotator's per-beat chord label. We are interested in the disagreement in the function in bar 45 (last beat) and bar 46.

3.4 Results: Statistics on Riemann

Overall, the majority (39%) of the disagreements (that Riemannian theory can explain) between the annotators' chord labels are substitutions of the tonic function. As found by Trevor De Clercq and David Temprely (2011) and Ashley Burgoyne (2012), the tonic (in their case just `I`, not including its possible substitutions) is often the most prominent chord in a popular music corpus, followed closely by the subdominant, and finally, the dominant.[53] This also aligns with Nicole Biamonte's work on the 'stable tonic,' 'less stable subdominant' and 'unstable dominant' as a way of generalising chord patterns in popular music.[54] There is a large difference between the number of chord-label disagreements that are explained by the major tonic (`T`, 29%) and the minor tonic (`t`, 10%). This substantial difference between the use of the major and the minor modes relates to the work of Trevor De Clercq and David Temprely (2011), who found the same dominance of the major mode in their Rolling Stones corpus.[55] The next most frequent Riemannian function found in our analyses is the subdominant function (`S` or `s`), which explains 34% of the chord label disagreements — again aligning with the prior work of De Clercq and Temperly (2011), Burgoyne (2012) and Biamonte (2010 and 2012).[56]

53. Clercq and Temperley, "A corpus analysis of rock harmony."; John Ashley Burgoyne, "Stochastic Processes and Database-Driven Musicology" (PhD diss., 2012).
54. Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music"; Biamonte, "Modal Function in Rock and Heavy Metal Music."
55. Clercq and Temperley, "A corpus analysis of rock harmony."
56. Clercq and Temperley, "A corpus analysis of rock harmony."; Burgoyne, "Stochastic Processes and Database-Driven Musicology"; Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music"; Biamonte, "Modal Function in Rock and Heavy Metal Music."

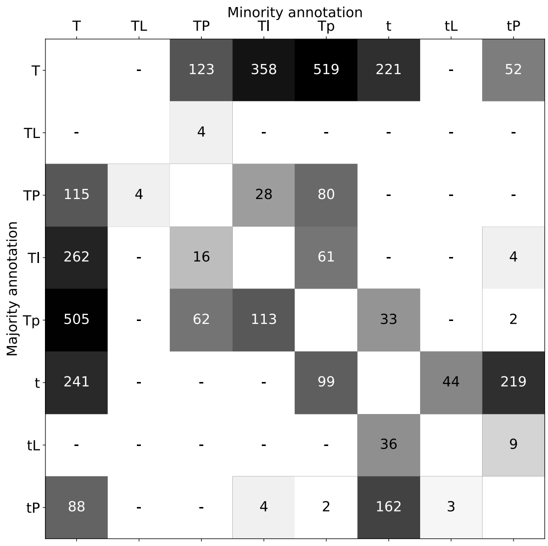


FIGURE 3.7: Frequencies of *tonic* substitution co-occurrences in the `Agreement` subset of the CASD. The numbers represent the frequency of co-occurrence of a majority substitution class (the most frequent substitution shared between the four annotators for a single beat) and other substitutions. The disagreements most often occur between *Parallele* — and *Variante* — related chords.

Like the tonic functions, most of the subdominant functions (28%) are of the major `S` function, while a much smaller percentage (6%) are explained by the minor `s`. Finally, 26% of the chord label disagreements can be explained through a dominant (`D` or `d`) function. Again, the major mode explains more disagreement, with (`D`) amounting to 17%, and the minor (`d`) explaining 9%.

Substitutions of the tonic function are the most commonly agreed upon chords in the `Agreement` subset (like the whole dataset).[57] Figure 3.7 shows the frequencies of tonic substitution co-occurrences in the `Agreement` subset, meaning the graph highlights which substitutions are likely to be perceived by the minority (X-axis) when the majority of the annotators agree on the substitution label on the Y-axis. The graph therefore reveals which substitutions most often appear in explaining the chord label disagreement between the annotators for the `Agreement` category. Within the tonic function, `T` often occurs with `Tp`, `Tp` also often occurs with `T`, and `T` with `Tl`. Therefore `T` and `Tp` are the most commonly confused chords. Therefore, there is a strong perceptual confusion between `T` and `Tp`.

Unsurprisingly, the most likely chords to be confused are the harmonic functions and single substitutions (e.g. `T` with `Tp` — chords that have two common tones, and are therefore very similar in their components). A small corpus of scholarly literature has focused on the similarities between a chord and its *Parallele*; one such work comes from

57. When discussing the results of this dataset, the focus of the discussion will be on aspects of each category that we can explain using Riemannian theory. Therefore, in the categories `Agreement` and `Partial Agreement` we will focus on the same function chord labels. Across-function disagreement will also be discussed; this will be concerning the discussion of the category `No Agreement`.

Carol Krumhansl, Jamshed Bharucha and Edward Kessler (1982), who found a pattern of correlations reflecting a strong relationship between a major scale and its relative minor.[58] The sheer prominence of the *Parallele* in European tonal art music for variation in musical forms, such as Theme and Variations and sonata form, also highlights the relationship of a key/chord and its *Parallele* as one that is similar enough to provide continuity within a change of harmony. The chord's two common tones also highlight their similarity, for example, between C major (C, E, G) and A minor (A, C, E). However, the idea that this similarity is perceived, or that this type of similarity could cause us to confuse chords that are related by a *Parallele* substitution, is speculation, and more research is required to evidence the audible similarities between chords related via this substitution. Interestingly, for chords related by a single substitution (e.g. the tonic, and its *Variante*, *Parallele* and *Leittonwechsel*), the mode of the substitution chord changes. However, literature often highlights the distinct differences between major and minor harmonies in terms of their characteristics,[59] and thus often places weight on the distance between these keys, and not their perceptual similarities. In this chapter, there are a vast number of instances where participants have disagreed on the mode. Thus, from an auditory point of view, there is a perceptual similarity between them. Lastly, Carol Krumhansl (1998) discusses the psychological reality of neo-Riemannian transformations (distinct from substitutions in terms of Riemannian theory — refer to Chapter 1 Section 1.3.2 for a discussion on neo-Riemannian theory).[60] She highlights that the perceptual similarity between chords related by the *Leittonswechsel* transformation is due to the importance of pitch proximity and the fact that it requires the shift of a single note by just one chromatic step — this is the same for our *Leittonswechsel* substitution, where the chord requires a single pitch shift, thus retaining two common notes.[61]

The most common subdominant functions to be disagreed upon can be seen in Figure 3.8. The chords most commonly disagreed upon are `S` with `Sp` and `Sp` with `S` — again showing the *Parallele* substitution to be the most important in explaining disagreements. Substantially less common is `S` and `Sl` (yet, still with greater frequency than other substitutions).

The most common dominant functions to be disagreed upon can be seen in Figure 3.9; `Dp` with `Dl`, `d` with `dP` and `D` with `Dp`. The high levels of confusion between `Dp` and `Dl` are surprising as, though they are both related (via the dominant function, both being substitutions of `D`), these chords only have one common note. Their confusion, therefore, may have more to do with their relationship to the dominant; it could be that the listener

58. Krumhansl, Bharucha, and Kessler, "Perceived Harmonic Structure of Chords in Three Related Musical Keys."

59. Marianna Pinchot Kastner and Robert G. Crowder, "Perception of the Major/Minor Distinction: IV. Emotional Connotations in You," *Music Perception* 8, no. 2 (1990): 189–202; Robert G. Crowder, "Perception of the Major/Minor Distinction: I. Historical and Theoretical Foundations," *Psychomusicology: Music, Mind, and Brain* 4, no. 1 (1984): 3–12.

60. Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations."

61. Ibid.

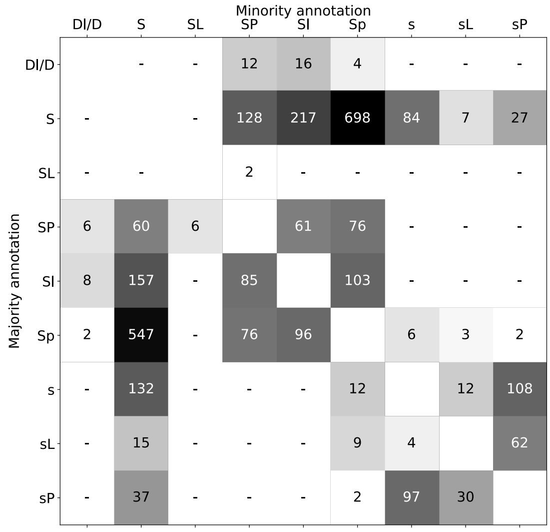


FIGURE 3.8: Frequencies of *subdominant* substitution co-occurrences in the `Agreement` subset of the CASD.

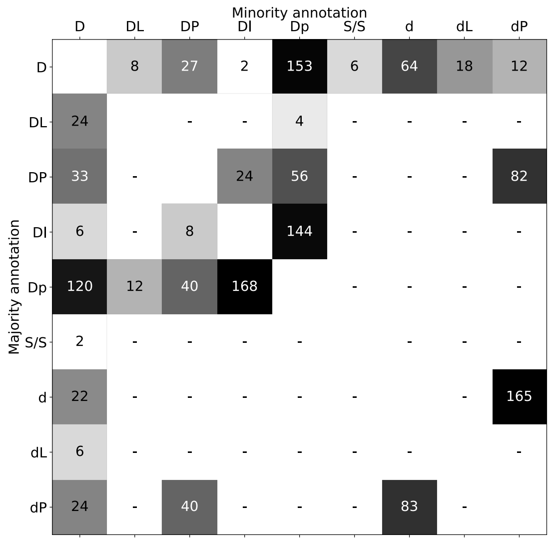


FIGURE 3.9: Frequencies of *dominant* substitution co-occurrences in the `Agreement` subset of the CASD.

hears a chord as having a dominant function rather than the pitches or specific chord, and therefore the listener hears a chord's function, rather than a specific collection of pitches. Also, similarly to the tonic and subdominant, the *Parallele* substitution is often useful for explaining disagreement. When exploring dominant function disagreement, we also found that the minor dominant was confused frequently with its major *Parallele*; something not found for either of the other functions.

The most common substitutions to be disagreed upon in the category `Partial Agreement` were `Sp` with `S`, `S` with `Sp` and `d` with `dP`, again highlighting the use of the *Parallele* substitution. Like the `Agreement` category, the `Partial Agreement` also uses the dominant minor mode. This use of the minor mode only seems to feature in reference to the dominant function.

Contrastingly, within the `No Agreement` category, where we can examine disagreements across functions, the most commonly occurring disagreements were found between `T` and `D`, followed by `S` with `T`, then `D` with `T`, `T` with `S`, and finally `S` with `D`. The most prominently occurring disagreements are between the basic functions, without any substitutions. The most common disagreements, therefore, seem to include the tonic function. Listeners hear chords as a function, rather than as a specific collection of pitches.

Using the music-theoretical method discussed in this chapter, at the function level, a total of 48% of the harmonic disagreements in CASD can be explained. Together with the sections of full agreement (66%) a little over 82% of the dataset is explained. Firstly, using Riemann's theory of harmonic functions in full (`Agreement`) explains 27% of the disagreement in this dataset, and a further 13% partially (`Partial Agreement`), totalling 40% explainability using this method. A further 5% is explainable through observing a disagreement over a chord caused by prolongation and granularity disagreements, such as the disagreement discussed for the song 'All Those Years Ago' by George Harrison. Finally, the score provides the opportunity to account for another 3% of the annotator disagreements, by highlighting ambiguities (for example the parts having different harmonies such as the disagreements discussed in Rick James' 'Super Freak').

3.5 Discussions and Conclusions

This chapter has presented a new application of Hugo Riemann's theory of harmonic functions as a method for explaining chord-label annotator disagreement. Using this approach, 48% of the harmonic disagreements in CASD can be explained. In full, Riemannian theory explains 27% of the disagreements between annotators and a further 13% partially. I supplemented this approach through utilising other information from the scores, enabling the explanation of a further 5% through disagreements caused by granularity, and another 3% through harmonic ambiguity. I have shown that music theory can explain some harmonic inter-annotator disagreement, demonstrating a higher level of agreement between annotators at this more musically informed harmonic function level.

The majority (40%) of the disagreements among the annotators' chord labels could be explained through substitutions of the tonic, followed by the subdominant (35%) and finally the dominant (25%). As discussed, these results align with previous work on popular music corpora and music-theoretical explorations of popular music harmony.[62]

62. Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music"; Biamonte, "Modal Function in Rock and Heavy Metal Music"; Clercq and Temperley, "A corpus analysis of rock harmony."; Burgoyne, "Stochastic Processes and Database-Driven Musicology."

Observing in detail the `Agreement` category, it was found that annotator disagreement was most frequently explained through a disagreement between a function and a single substitution (e.g. `T` with `Tp`), except in the case of `Dp` and `Dl`. The *Parallele* substitution was the most frequent substitution to feature as an explanation of harmonic disagreement — this being the 'relative' relationship. I showed that this is likely to be because chords related by one substitution will have two common notes, since they are related by a single pitch shift.[63] However, the idea that, therefore, we perceptually hear a similarity between these chords related by one substitution is currently speculative. The results of this chapter provide an impetus, warranting the exploration of Riemann's theory's relationship to music perception.

As previously discussed, the metrics used to measure annotator disagreement in MIR studies commonly focus on pitch-class agreement. These methods paint too bleak a picture of the agreement between annotators. For example, the chord labels `C:sus4` and `A:min` have no root note agreement, and no agreement on the root and third using the common MIREX evaluation measures. However, when analysing these chords in the key of `C`, Riemannian theory reveals that they both fulfil a tonic function (as `T` and `Tp`, respectively). An initial analysis of CASD shows that within the part of the dataset that is fully explainable using our music-theoretical approach (`Agreement`), there is only approximately 49% root note agreement, and even less (39%) agreement on the root and the third. Therefore, there is a large difference in the notion of chordal agreement between the two approaches. A large scale comparison of pitch-class oriented methods with the function-oriented music-theoretical approach could reveal how to inform the current evaluation methods in MIR. In turn, this would enable the creation of metrics that take into account the function of a chord in a tonal centre, providing a more nuanced view on chordal agreement and similarity. Future work should look into whether music-theoretical approaches for explaining inter-annotator disagreement are useful for explaining inter-annotator disagreement in datasets which include non-musically trained individuals (such as crowd-sourced harmony datasets). It would be worth exploring whether this academic way of thinking about harmony is related to the general population, or is only applicable to those who have musical training. In turn, the larger number of annotators available through crowd-sourcing could enable statistically relevant results. I will explore using a crowd-sourced dataset in Chapter 4.

It is worth noting the limitations of this study. This analysis was completed on a dataset containing diverse popular music and annotators, but pales in comparison to the number of transcriptions found in online repositories, which raises questions on the ability to generalise the results. A larger dataset could provide more insight into factors (e.g. primary instrument) that influence chord label choice, and an empirical upper limit of inter-annotator agreement of harmonic function. However, creating a large enough

63. Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations."

dataset to investigate these properties with statistical validity is time-consuming and costly. Chapter 4 provides an analysis of a larger dataset through collating harmonic annotations made through crowd-sourcing. This chapter's two studies shows that the methodology of this study is also helpful in explaining annotator disagreement between a more significant number of annotators. Though this sample size is limited, this research however grows our understanding through a more qualitative exploration of Annotator disagreement. This work provides a methodological novelty for both MIR and Music Theory, providing an impetus for further work in this area.

It is also important to note that the methodology of this chapter requires a musical score to perform the Riemannian analysis (to enable the determination of any key changes within the music). Requiring a score led to the removal of 9 songs from the dataset, as they had no available score. Due to the recording and compositional practices of popular music, scores are often only available through (subjective) transcription. Chapter 5 will provide an analytical approach that removes the need for a score by removing the functional element from Riemann's theory. These nine songs, and the two songs from Chapter 4, which also did not have available scores, will be discussed and analysed.

This chapter has shown that some assumed annotator disagreement is a form of agreement (or perceptual similarity) on a functional level, which results in a more nuanced view of inter-annotator disagreement. Showing which chords are perceived to be similar is important for the study of music similarity and harmonic similarity in particular. These results should inform future similarity measures used in music similarity tasks and computational harmony tasks such as ACE by taking into account the function of a chord, instead of merely its pitch-class makeup. With the growing number of studies into annotator disagreement, computational harmony analysis will inevitably move towards modelling the perceived (or subjective) harmony of multiple annotators.

Chapter 4

Using harmonic theory to explain inter-annotator disagreement in crowd-sourced repositories

4.1 Introduction

A vast number of harmonic annotations exist in crowdsourced repositories (e.g., Ultimate-Guitar,[1] and Chordify).[2] They provide a valuable resource for the exploration of harmonic inter-annotator disagreement between a larger number of annotators than previous research (for example, larger than the previous chapter's dataset). Crowdsourced repositories employ the productive potential of millions of enthusiasts brought together by the World Wide Web.[3] Crowdsourcing as a phenomenon has received increasing attention in academic research since Jeff Howe coined the concept in 2006.[4] Howe defined crowdsourcing as 'the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call'.[5] Popular examples of the numerous systems that have utilised a crowd to gather data and solve a problem include: Wikipedia,[6] SETI@home,[7] and Open Streets Map.[8]

Crowdsourcing has seen increased application in ever more complicated tasks, causing debate around how one can control the quality of this data. Some crowdsourced tasks

1. https://www.ultimate-guitar.com/
2. https://chordify.net/
3. Jeff Howe, "The Rise of Crowdsourcing," *Wired magazine* 14, no. 6 (2006): 1–4; Anhai Doan, Raghu Ramakrishnan, and Alon Y. Halevy, "Crowdsourcing systems on the world-wide web," *Communications of the ACM* 54, no. 4 (2011): 86–96.
4. Howe, "The Rise of Crowdsourcing."
5. Ibid.
6. https://en.wikipedia.org
7. https://setiathome.berkeley.edu/
8. https://www.openstreetmap.org/

can be complicated; for example, translating text requires specific domain knowledge of both languages involved in the translation. The quality of crowdsourced data cannot be guaranteed.[9] Furthermore, Yuxiang Zhao and Qinghua Zhu (2014) noted that it could be an issue just to find participants in the first place, regardless of their appropriateness for the task.[10] Osamuyimen Stewart et al. (2010) highlighted the risk of the participants in crowdsourced work completing only part of the work, leaving an incomplete dataset, or unfinished data.[11] Researchers have therefore proposed methodologies for both maintaining a crowd's interest throughout an activity,[12] and for selecting a high-quality crowd that meets your requirements for the task at hand.[13]

In the field of music, we have seen the potential of utilising the crowd to complete a creative process. One example is crowdsourced composition, where individuals all over the world (with Internet access) can contribute to the process of music production, although they may not be professionals or experts.[14] Crowdsourcing in music has already been successful, and usually features expert oversight. Projects, such as the DJ Avicii 2013 song 'X You' which utilised crowdsourced sounds, Eric Whiticar's crowdsourced 'virtual choir', the collective work of Detroit Michigan's City Orchestra and Tod Machover, and Maroon 5 fans, who helped to compose a song in 24 hours.[15] In this Maroon 5 collaboration, fans were called upon to recommend lyrics, riffs, and rhythms through social networking services such as Facebook and Twitter, leading to the composition of the song 'Is Anybody Out There?'

The projects detailed so far have featured experts tasked with overseeing the creative process, and therefore influencing the outcome. Brendon Feris had the idea to bring together the collaborative potential of individuals through the website *Crowdsound*, to write a song collectively.[16] Feris built a simple music player and, although he did not directly compose the song, he placed some rules to restrict its development. He began the melody with a C followed by a D, and then built a voting system that asked visitors to the site in real-time to vote on the next note in the sequence.[17] To control the melody,

9. Christoph Riedl et al., "Rating Scales for Collective Intelligence in Innovation Communities: Why Quick and Easy Decision Making does not get it Right.," In *Proceedings of 2010 International Conference on Information Systems* (St Louis), 2010,

10. Yuxiang Zhao and Qinghua Zhu, "Evaluation on Crowdsourcing Research: Current Status and Future Direction," *Information Systems Frontiers* 16, no. 3 (2014): 417–434.

11. Osamuyimen Stewart, Juan M. Huerta, and Melissa Sader, "Designing Crowdsourcing Community for the Enterprise," In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 2009, 50–53.

12. Ibid.

13. Vikas C. Raykar and Shipeng Yu, "Eliminating Spammers and Ranking Annotators for Crowdsourced Labelling Tasks," *Journal of Machine Learning Research* 13 (2012): 491–518.

14. Carlos Gomes et al., "Crowdsourcing for Music: Survey and Taxonomy," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on* (IEEE, 2012), 832–839.

15. Todd Wasserman, *Coca-Cola To Help Maroon 5 Crowdsource a New Song*, Available at: https://mashable.com/2011/03/01/coca-cola-maroon-5, March 2011; Aviva Rutkin, *Crowdsourced Song lets the Masses Compose — One Note at a Time*, Available at: https://www.newscientist.com/article/dn28105-crowdsourced-song-lets-the-masses-compose-one-note-at-a-time/, August 2015.

16. https://crowdsound.net

17. Rutkin, *Crowdsourced Song lets the Masses Compose — One Note at a Time*.

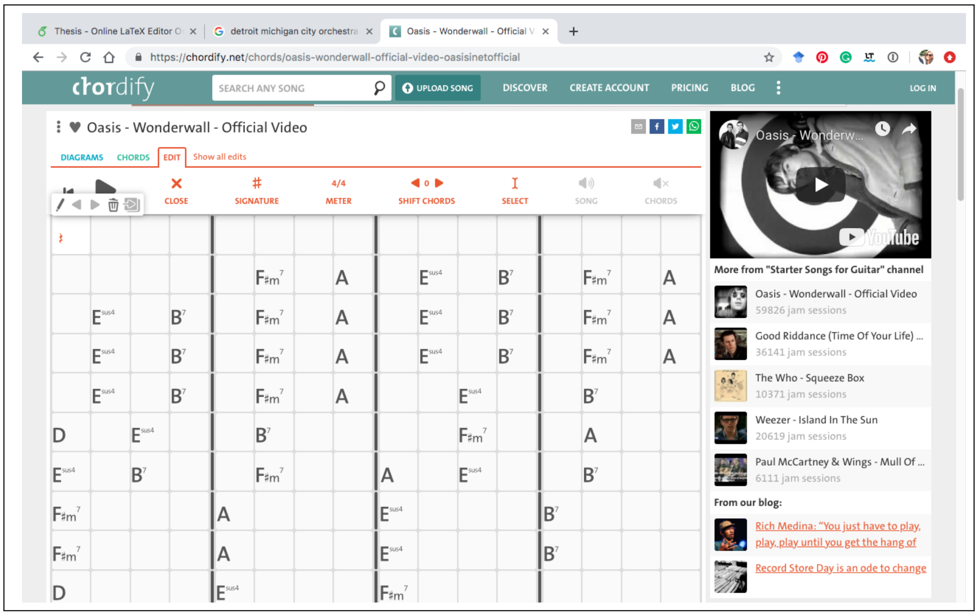


FIGURE 4.1: The chord edit page for the song ‘Wonderwall’ by Oasis, on Chordify.net accessed 23rd April 2019

Ferris chose the chord progression of C, G, Am, F (`I, V, vi, IV`) to repeat throughout the piece — this is a widespread popular music progression (for example it is used in ‘Africa’ by Toto, ‘Photograph’ by Ed Sheeran, and ‘The Scientist’ by Coldplay). He also set the length of the song to 18 repetitions of this sequence, and stipulated the structure verse, verse, chorus, verse, chorus, conclusion. In total, the song had 67,167 votes involved in its creation. 50 votes were required for each pitch before the system selected the most popular and moved on to the next.[18]

The World Wide Web has enabled individuals to post their annotations of lyrics and chords for others to view, amend and edit. These annotations are increasingly collated in the development of websites that utilise crowdsourced guitar chords and guitar tablature — such as the Ultimate Guitar Repository and Chordify. Chordify did not begin as a crowdsourced harmonic annotation website; instead, Chordify describes itself as ‘an online music education service — made for and by music enthusiasts — that transforms music from Youtube, Deezer, SoundCloud, or your private collection into chords’.[19] The service recognises the chords of a piece from an audio signal, and aligns them on a beat tracking grid to allow users to play along to their chosen tracks. In 2015, Chordify introduced the ‘chord edits’ feature,[20] allowing their users to edit the chord outputs they generated automatically, and other users’ edits, through an interactive interface (see Figure 4.1 for a screenshot of this interface), introducing a crowdsourced element. By aligning the edits made by Chordify users, we can create a dataset of harmonic disagreement.

18. Ibid.
19. see https://chordify.net/pages/about/ for more information
20. See Chordify’s post on their feature: https://chordify.net/pages/how-to-use-chordify/chord-edits/

The study of harmonic inter-annotator disagreement is yet to utilise crowdsourced repositories. In the previous chapter, I explored the possibility of using Hugo Riemann's theory of harmonic functions to explain annotator disagreement. This chapter showed that for CASD Riemannian theory could explain 40% of the disagreement in the dataset, a further 5% through observing a disagreement over a chord caused by annotator granularity, and a further 3% through examination of the score and observation of ambiguities that are present within the music. Chapter 3 raised concerns over the size of the dataset, having only four annotators. Therefore, this chapter will continue to explore annotator subjectivity through Chordify's crowdsourced chord label annotations, using a dataset of 77 different annotators, with up to 11 annotators per song.

This chapter will begin with an analysis of a subset of Chordify's user-edit data in the same manner as Chapter 3. Following this, a qualitative interview study will explore why and how individuals made chord edits. Overall, this chapter will demonstrate that Riemannian theory helps explain some annotator disagreement. I will show that Riemannian theory explains as much as 50% of the disagreements in this chapter's dataset. Disagreements present in the score explain a further 3% of the annotators' disagreements (such as prolongation, granularity disagreement and musically explained disagreements), and a further 15% of annotator disagreement arises from disagreement over the exact placement of a chord change. In total, I explain 68% of the disagreements in this dataset. The conclusion proposes that our ability to explain a more substantial amount of harmonic disagreement may be due to amateur annotators having a smaller range of musical vocabulary, leading to less disagreement. A particular noteworthy result relates to the aims of Participant 3 in the second study of this chapter: they discussed their use of Riemannian-based approaches (though they did not know the approach's name) to simplify chords for beginners. This acknowledgement of Riemannian theory, along with specific mentions of granularity annotation decisions, highlights the involvement of this music-theoretical approach (defined in this chapter and Chapter 3) in the perception of harmonic similarity. Overall, this chapter highlights that Riemannian theory is a suitable tool to assess musical similarity.

4.1.1 Dataset

This chapter uses a subset of the approximately 580,000 user edits made on Chordify between 2016 and 2018, using the 'chord edits' function. Due to the sheer number of edits created (and the accompanying textual data), Chordify reduced this dataset (for their research) to 11,638 edits. They did so by through filtering based on three rules: (1) the edits had to be created on a recent algorithmic output, to ensure the quality of the original labels the users had to edit; (2) the users had to have edited at least ten beats to ensure that they were actively editing the songs, not just trying out the feature; and (3) the annotators had to have edited multiple songs to ensure they had some basic experience in transcribing harmony. The outcome of this filtering resulted in the 11,638–edit subset.

For the purposes of this study, I reduced these 11,638 edits further through the enforcement of the additional rules below. I used these rules to ensure that the dataset produced was suitable for a Riemannian function analysis, and of a meaningful, but manageable size for the analytical processes that needed completion by hand. The rules stipulated were that the edit had to:

1. Have a working YouTube URL.
2. Have edits that span the whole song.
3. Have edits that were more complex than just changing the timing of a chord.
4. Have edits that were more complex than just adding sustained notes.
5. Be from a song longer than 160 beats.
6. Be from a song that had edits made by at least three users (and the original annotation).

The first rule ensured that the original YouTube video of the song, used to create the edits, was still accessible. I needed these videos both for the identification of the song title and artist's name, along with the audio file enabling the manual alignment of the annotators' chord edits with the scores. The second rule aimed to ensure as far as possible that the edits were made purposefully, and were not just the user testing out the function (similar to the second rule applied during Chordify's initial filtering stage). I assumed that a user who changes chords throughout the song was more likely to have done this purposefully. The third and fourth rules reduced the dataset to songs that feature disagreement on the fundamental harmony of the piece. The primary purpose of this was to ensure that the harmonic changes were substantial enough to use Riemannian theory to account for these changes; Section 3.1.3 of the previous chapter discusses this further. Fundamentally, the harmony had to be changed and not just decorated with sustained notes. The fifth rule was implemented because I observed that songs shorter than 160 beats were jingles or commercials that users wanted to play along with, and not necessarily a piece with harmony; more likely they were just a melodic line (and unlikely to have an available score transcription). The final rule ensured that the number of annotations per song (four including the original transcription) is at least as large as previously created datasets — such as the one analysed in Chapter 3. Therefore, this chapter enables the observation of whether the methodology defined in the previous chapter (Section 3.2) applies to other datasets with the same and more significant numbers of annotators.

A total of 148 edits, from 41 different songs by 77 annotators resulted from this filtering process and used in the subsequent analysis.

4.1.2 Disagreement

This chapter uses a global agreement analysis to highlight sections of overall disagreement, as defined in Section 3.1.3 of Chapter 3. This methodology assumes that when an annotator did not make an edit to a chord label, they agreed with the given chord label for that segment. Overall, there is an 18% disagreement between the annotators in this chapter's dataset. In comparison, the previous chapter found a 34% disagreement. Like the previous chapter, the levels of disagreement will be the focus of this chapter.

4.2 Study 1: Analysis of User Edits

As explained in the introduction (Section 4.1), this chapter features two studies. The first takes a similar methodological approach to that of the previous chapter (Chapter 3 Section 3.2), the only changes were in the alignment stages as detailed below.

4.2.1 Methodology

As in Chapter 3, the harmonic annotations were aligned with the musical scores to enable a Riemannian analysis. The annotations also needed to be aligned with each other in this dataset (before their alignment to the scores).

Chordify adjusted their beat-tracking algorithm between 2016 and 2018, meaning that the edits for this dataset may have been based on different beat segments or even different time signatures. Therefore, the different harmonic annotations could not simply be matched with each other, as slight differences in beat detection could result in a false disagreement. For example, if one beat tracker had perceived a song in three beats per bar, and another tracker perceived it as in four beats per bar, aligning the beats would not match the bar lines accurately (see Table 4.1). Therefore, a reference annotation was chosen randomly from one of the beat annotations, and I then aligned the rest of the annotations to it by observing the time indices of the reference annotation and finding the corresponding chord labels for the other annotators (see Table 4.1). This method could potentially result in a bias towards a particular beat tracker. However, this effect was controlled by manually checking and correcting the alignment of the beat.

Following the beat alignment of the annotations, each aligned 'beat' was then aligned with the beats and bar numbers of the score, using the same process detailed in Chapter 3, Section 3.1.1. Sheet music was successfully sourced (from *Musicnotes*) for 39 out of the 41 songs. Thus, the dataset was reduced to these 39 songs (Chapter 5 discusses the two songs that have no available scores in this chapter, along with the nine from Chapter 3).

A1	Bar 1 Beat 1 0:00-0:10 C	Bar 1 Beat 2 0:10-0:20 G	Bar 1 Beat 3 0:20-0:30 F	Bar 1 Beat 4 0:30-0:40 A
A2	Bar 1 Beat 1 0:00-0:13 C	Bar 1 Beat 2 0:13-0:26 G	Bar 1 Beat 3 0:26-0:39 F	Bar 2 Beat 1 0:39-0:52 C

	0:00-00:13	0:13-0:26	0:26-0:39
A1	Bar 1 Bar 1 (C)	Bar 1 Beat 2 (G)	Bar 1 Beat 3 (A)
A2	Bar 1 Beat 1 (C)	Bar 1 Beat 2 (G)	Bar 1 Beat 3 (F)

Table 4.1: Comparisons between two possible beat tracking algorithms used by Chordify. The first row shows a beat tracker extracting 3/4 and the second row the same chord sequence extracted in 4/4. This table aligns the segments as if labelled segment 1, 2, 3, 4 etc. This shows how bar 2 beat 1 of the 3/4 time signature, would align with bar 1 beat 4 of the 4/4 time signature.

The remainder of this chapter's methodology follows that which is detailed in Section 3.2 of Chapter 3. I began by ascertaining the substitution label for each chord-label for each annotator of a song (according to Riemannian theory). Next, all annotators' substitutions were compared for a single beat. Those beats that had some annotator disagreement (the 18% detailed in Section 4.1.2) were then categorised as either `Agreement`, `Partial Agreement` or `No Agreement` with respect to harmonic functions.

After this, further examination of the scores allowed us to see if they could explain any of the remaining annotator disagreements, including disagreements in the harmony caused by different instruments playing different chords and disagreement on the level of granularity at which to annotate (Section 3.2.1 in Chapter 3 provides further detail on this disagreement). The different beat-tracking algorithms, and the alignment process required, could have also lead to disagreement over the precise segment in which the harmonic change occurs; to identify, for example, cases where all annotators agree that the music changes from C major to G major but A_1 and A_2 perceive the change happening on beat 1 of the bar, and A_3 and A_4 on beat 2. I observe this as an explainable disagreement, as the annotators agree on the fundamental harmonic change. Also, I acknowledge that the beat alignment process may have caused this disagreement.

4.2.2 Results: Example Analyses

This section will discuss four example analyses from the subset of Chordify user edit data including 'Take on me' by A-ha, 'Hotel California' by The Eagles, 'Africa' by Toto, and 'Over the Rainbow' by Israel 'Iz' Kamakawiwo'ole. The examples show a mixture of the different categories `Agreement`, `Partial Agreement`, and `No Agreement`. Each example will begin with a discussion highlighting the ability of Riemann's theory of harmonic functions to improve our understanding of disagreement; following this any remaining disagreements will be explained through features in the score, and the segment in which the harmonic change occurs.

4.2.2.1 'Take on Me' by A-ha

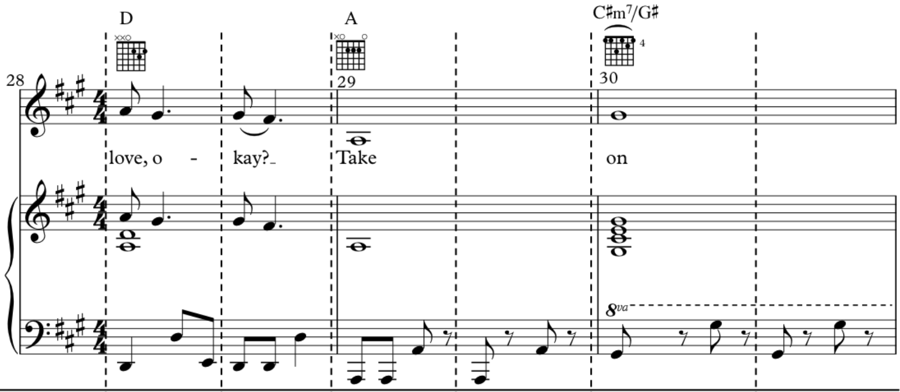
A-ha's 'Take on Me', in A major, had chord-labels edited by five annotators. There is a 41% disagreement between the annotators for the harmony in this song. 57% of the annotator disagreement in this song is explained in full using Riemannian theory, and 10% partially.

Bars 28–33 (Figure 4.2) shows the disagreement present on the title lyrics 'Take on Me'. 29% of the harmonic disagreements in A-ha's song take place during this phrase. The melodic line associated with the opening chorus lyrics 'Take on me', (bars 29–31) rises an octave from the pitch A3 to an A4, passing through the pitch G♯. The annotators disagree on the level of granularity at which to annotate the harmony. Annotators A_2, A_3 and A_5 perceived the G♯ as a passing tone and therefore harmonically unimportant. In turn, there is a prolongation of the harmony of A major (`T`) throughout the three bars. In contrast, the other annotators perceive the harmony as changing more frequently.

The beginning of bar 30 shows an example of the category `Partial Agreement`: annotators A_2, A_3, A_4 and A_5 agree on the tonic function, with A_2, A_3, and A_5 agreeing on A major (`T`) and A_4 perceiving it as C♯ minor (`Tl`). The one annotator that disagrees with the function is A_1, who perceives E major (`D`). Musically, this complex chord features the pitches of both the E major and C♯ minor chords (except the B (the 5th of E major)).

The last segment, of bar 30, is an example of the category `Agreement` — where all five annotators agree on the tonic function. Specifically, with A_1 perceiving the chord as F♯ minor (`Tp`), A_2, A_3, and A_5 as A major (`T`) and A_4 as C♯ minor (`Tl`). A_1's choice of F♯ minor could be explained by the bar that follows, where they also perceive the chord F♯ minor. Therefore, the beat alignment phase of this chapter's methodology has likely caused this misalignment.

Bar 31 also presents examples of the categories of `Agreement` and `Partial Agreement`. The first beat of bar 31 is the category `Agreement`, as the annotators agree on the tonic function. A_2, A_3 and A_5 agree on the chord A major (`T`), and annotators A_1 and A_4 agree on F♯ minor (`Tp`). The second segment (beats 3 and 4) of bar 31 provides an example of the category `Partial Agreement`, as A_2, A_3, A_4 and A_5 agree on the tonic function. Whereas, the remaining annotator (A_1) annotated the subdominant function (D major), which aligns with the transcription notated in the score. Again, this example suggests a possible beat alignment issue.



A_1	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)	E:maj (D)	F#:min (Tp)
A_2	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)	A:maj (T)	A:maj (T)
A_3	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)	A:maj (T)	A:maj (T)
A_4	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)	C#:min (Tl)	C#:min (Tl)
A_5	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)	A:maj (T)	A:maj (T)
Disagreement category					P	A

A_1	F#:min (Tp)	D:maj (S)	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)
A_2	A:maj (T)	A:maj (T)	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)
A_3	A:maj (T)	A:maj (T)	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)
A_4	F#:min (Tp)	F#:min (Tp)	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)
A_5	A:maj (T)	A:maj (T)	D:maj (S)	D:maj (S)	A:maj (T)	A:maj (T)
Disagreement Category	A	P				

FIGURE 4.2: Bars 28 to 33 of 'Take on me' by A-ha. The figure shows the musical score aligned with each of the five annotators' per-beat chord label. We are particularly interested in the disagreements present for the title words 'Take on me'.

Though the pitches change for the lyrics 'Take on Me' later in the song (for example bars 71–73 the melodic line raises from a C♯ to an A), disagreement continues on these title lyrics. The disagreement in terms of the hierarchical level at which to perceive the harmony explains 5% of the disagreements between annotators within this song. In total, this means a combination of this method and Riemannian theory can explain 72% of the disagreements in this song.

4.2.2.2 'Hotel California' by The Eagles

The second example comes from The Eagles' 'Hotel California', which had 11 annotators — the largest number of annotators for a song in this dataset. There was a 12% disagreement between annotators, which considering the number of annotators means there is a strong level of agreement (88%). For this song, we can explain 48% of this disagreement entirely using Riemannian theory, and 2% partially (leaving 50% unexplained disagreement).

The home key is B minor. Bars 55 to 59 (Figure 4.3) feature multiple instances of the category `Agreement`, including both segments in bar 56 (agreeing on the tonic function), and the first segment of bar 59 (also agreeing on the tonic function). The first example from bar 56 shows annotators A_9 and A_{11} disagreeing with the other annotators who perceive B minor (`t`), and instead perceive it as D major (`tP`). The bass line may explain this disagreement, as it rises to a D at the start of bar 56, from the B at bar 55 (passing through a C♯). A_9 and A_{11}, therefore, may be annotating the harmony with the changing bassline and perceiving the D major chord as an important harmonic movement. As D major shares both the pitches F♯ and D with B minor, A_1, A_2, A_3, A_5, A_6, A_7, A_8 and A_{10} may instead be focusing on the held B minor right-hand chord, and do not perceive the bass line as changing the fundamental harmony.

Bar 59 beat 1 features annotator disagreement between the chords D major (A_1, A_3, A_5, A_7, A_9, A_{10} and A_{11}) and G major (A_2, A_4, A_6 and A_8). Though this disagreement is explained by Riemannian theory, as both chords are substitutions of the tonic (D major as `tP`, and G major as `tL`), the disagreement could also arise from the different beat-tracking algorithms. As all the annotators agree on the change between G major and D major, the disagreement is purely on which segment the chord change occurs in. Their disagreement may also have arisen from the syncopated rhythm of the melodic line which could distort the exact point of change of the harmony, as the melody line features the pitch G suspended into the start of bar 59. Therefore, multiple factors could have influenced this disagreement.

22% of the disagreement in this song is caused by annotators perceiving chord changes as occurring in different segments. In total, we can explain 74% of the disagreements in this song through both Riemannian theory (58%) and disagreement over where a harmonic change occurs (22%).

4.2.2.3 'Africa' by Toto

The third example is taken from Bars 1 to 5 of 'Africa' by Toto. This piece had five annotators. There is a 26% disagreement between the annotators, of which we can explain 37% using Riemannian theory, and 7% partially using Riemannian theory.

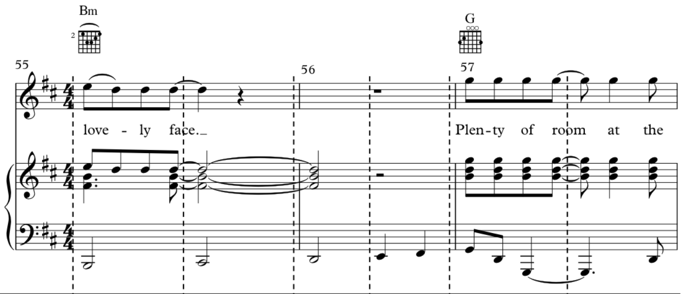
The home key is B major. Figure 4.4 features the famous piano riff that starts the piece and continues throughout. This piano riff provides an example of annotator disagreement that Riemannian theory cannot explain. The disagreement falls on beat 4 of bar 1, 3, and 5 (and inconsistently on beat 3 of bar 5). The annotators A_3 and A_5 (and inconsistently A_2) perceive the continuation of A major (`dP`), the same chord present for the rest of the bar. In contrast, A_1 and A_4 (and inconsistently A_2) perceive the chord as changing to C♯ minor (`Sp`) (the same as perceived by the transcriber who notated the score). These chords have different functions: the dominant (`dP`) and the subdominant (`Sp`). Observing the music, one can appreciate the uncertainty as to whether the chord changes at the last beat of the first bar of the two-bar phrase (e.g. bars 1, 3 and 5), or the first beat of the second bar of the phrase (e.g. bars 2, 4, and 6), due to the rhythmic tie in the melodic part and inner voices vs the downbeat articulation of the bass line. The lack of a G♯ or A in the held chord of bars 1–2 and 3–4 may also have caused this disagreement, as pitches C♯ and E could be a continuation of A major without its root, or C♯ minor without its fifth.

Interestingly, the chords A major (A, C♯, E) and C♯ minor (C♯, E, G♯) are themselves related by a *Leittonwechsel* transformation (the roots of the chords are a major third apart), but within the key of B major they do not hold the same function. This prominent piano riff is a cause for many disagreements in this song. We can explain 25% of this song's disagreement through a disagreement between A major and C♯ minor in the piano part. In total, 69% of the annotator disagreement in this song is explainable using a combination of Riemannian theory and the explanation of the piano riff disagreement.

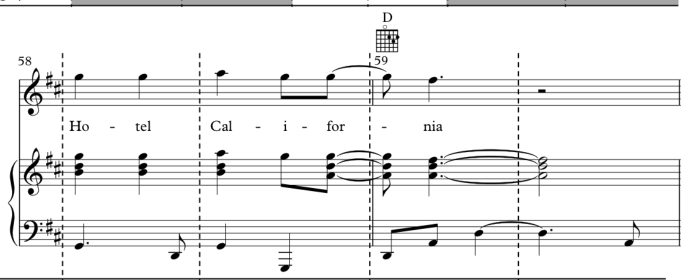
4.2.2.4 'Over the Rainbow' by Israel 'IZ' Kamakawiwo'ole

The final song from this dataset that I would like to discuss is 'Over the Rainbow' by Israel 'IZ' Kamakawiwo'ole. This piece had edits by seven annotators, with 31% disagreement between them. For this song, using Riemannian theory, I explain 41% of the disagreements fully, and 4% partially.

Figure 4.5 shows verse two, where bars 14 to 16 accompany the lyrics 'Somewhere over the rainbow, bluebirds fly'. This example shows multiple segments of disagreement.

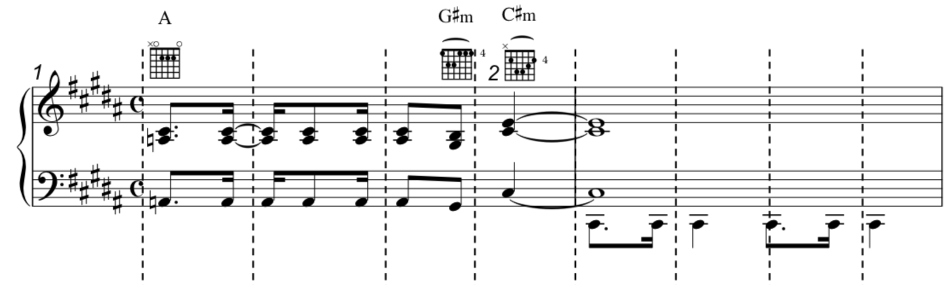


A_1	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_2	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_3	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_4	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_5	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_6	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_7	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_8	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_9	B:min (t)	B:min (t)	D:maj (tP)	D:maj (tP)	G:maj (tL)	G:maj (tL)
A_{10}	B:min (t)	B:min (t)	B:min (t)	B:min (t)	G:maj (tL)	G:maj (tL)
A_{11}	B:min (t)	B:min (t)	D:maj (tP)	D:maj (tP)	G:maj (tL)	G:maj (tL)
Disagreement Category			A	A		

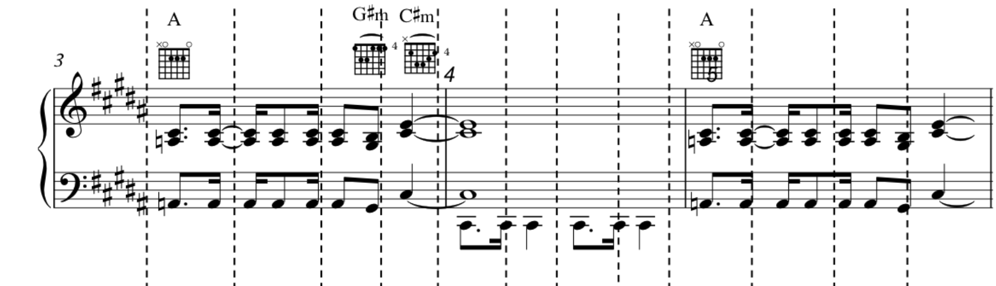


A_1	G:maj (tL)	G:maj (tL)	D:maj (tP)	D:maj (tP)
A_2	G:maj (tL)	G:maj (tL)	G:maj (tL)	D:maj (tP)
A_3	G:maj (tL)	G:maj (tL)	D:maj (tP)	D:maj (tP)
A_4	G:maj (tL)	G:maj (tL)	G:maj (tL)	D:maj (tP)
A_5	G:maj (tL)	G:maj (tL)	D:maj (tP)	D:maj (tP)
A_6	G:maj (tL)	G:maj (tL)	G:maj (tL)	D:maj (tP)
A_7	G:maj (tL)	G:maj (tL)	D:maj (tP)	D:maj (tP)
A_8	G:maj (tL)	G:maj (tL)	G:maj (tL)	D:maj (tP)
A_9	G:maj (tL)	G:maj (tL)	D:maj (tP)	D:maj (tP)
A_{10}	G:maj (tL)	G:maj (tL)	D:maj (tP)	D:maj (tP)
A_{11}	G:maj (tL)	G:maj (tL)	D:maj (tP)	D:maj (tP)
Disagreement Cateogry			A	

FIGURE 4.3: Bars 55 to 59 of 'Hotel California' by The Eagles. The figure shows the musical score aligned with each of the 11 annotators' per-beat chord label. We are particularly interested in the disagreements that happen at the beginnings and ends of bars.



A_1	A:maj (dP)	A:maj (dP)	A:maj (dP)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)
A_2	A:maj (dP)	A:maj (dP)	A:maj (dP)	A:maj (dP)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)
A_3	A:maj (dP)	A:maj (dP)	A:maj (dP)	A:maj (dP)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)
A_4	A:maj (dP)	A:maj (dP)	A:maj (dP)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)
A_5	A:maj (dP)	A:maj (dP)	A:maj (dP)	A:maj (dP)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)	C#:min (Sp)
Disagreement Category				N				



A_1	A:maj (dP)	*	*	C#: min (Sp)	*	*	*	*	A:maj (dP)	*	C#:min (Sp)	C#:min (Sp)
A_2	A:maj (dP)	*	*	C#: min (Sp)	*	*	*	*	A:maj (dP)	*	A:maj (dP)	C#:min (Sp)
A_3	A:maj (dP)	*	*	*	C#:mi n (Sp)	*	*	*	A:maj (dP)	*	A:maj (dP)	A:maj (dP)
A_4	A:maj (dP)	*	*	C#: min (Sp)	*	*	*	*	A:maj (dP)	*	C#:min (Sp)	C#:min (Sp)
A_5	A:maj (dP)	*	*	*	C#:mi n (Sp)	*	*	*	A:maj (dP)	*	A:maj (dP)	A:maj (dP)
Disagreement Category				N								

FIGURE 4.4: Bars 1 to 5 of 'Africa' by Toto. The figure shows the musical score aligned with each of the five annotators per-beat chord label. The symbol '*' means continuing the same harmony.

Throughout 'Over the Rainbow', we find most of the annotator disagreement is over where the chords start and finish (62%). In bar 16, the chord change that falls at the start/end of the bar is unclear. There is an agreement on the chord being F major (`S`) for the second segment of bar 16. Yet, the first segment has a disagreement on the harmony, with A_4 and A_7 perceiving G major (`D`), A_5 perceiving E minor (`Dp`) and A_1, A_2, A_3 and A_6 perceiving F major (`S`). Again, this was likely a result of aligning the segments. In this example, there are many segments (8 per bar), and the original beat-tracking algorithms perceived a differing number of segments for the song, ranging from 327–510. Therefore, when aligning the different algorithms, distortion is likely to occur at the beginning and end of the bar. Also, dependent on whether the user had greater or fewer segments to annotate, this may have changed the granularity at which they were able to annotate the harmony. We can see that annotators A_4, A_5, and A_7 likely had fewer segments to annotate. A_4, A_5, and A_7, throughout the extract, change harmony one or two segments later than the other annotators (see bar 14 segments 1–3, and bar 15 segments 1–3). The disagreement in bar 16 is therefore not specifically between E minor, G major and F major, but a disagreement on when the harmony of bar 16 changes to F major.

A lot of the disagreement in this song (62% as previously mentioned), falls at points of harmonic change (some of which can be explained by Riemannian theory). The fact that a large proportion of this song's disagreement is explainable using this method implies that the segmentation may not fall precisely on the beats, or it is unclear audibly when new harmony starts. This example reflects a common theme in this dataset, where multiple songs have proportions of disagreement that are explainable because they occur at points of harmonic change. These include 43% of the disagreement in 'Nikita' by Elton John and 39% for 'Better When I'm Dancin' by Meghan Trainor.

4.2.3 Results

This chapter has investigated whether Riemannian theory can explain some of the annotator disagreement in a subset of Chordify's user edit data. Figure 4.6 shows the percentage of the disagreement, in each song, that falls into each of the defined categories `Agreement`, `Partial Agreement` and `No Agreement`. The category `Agreement` explains 48% of the annotators' disagreements, 2% can be explained by the `Partial Agreement` category, and the `No Agreement` category explains 50% (see Figure 4.6 for a graph of each song against each of the three categories). In other words, this means Riemannian theory can explain 50% of the annotators' disagreements, at least partially.

The category `Partial Agreement` explains very little of the disagreement in this dataset; only 2% (shown by Figure 4.6 where the songs group closest to the `Agreement` axes). The song that had the largest percentage of explainable disagreement in the category `Agreement` was the song 'Under the Milky Way' by Church (18% disagreement). In

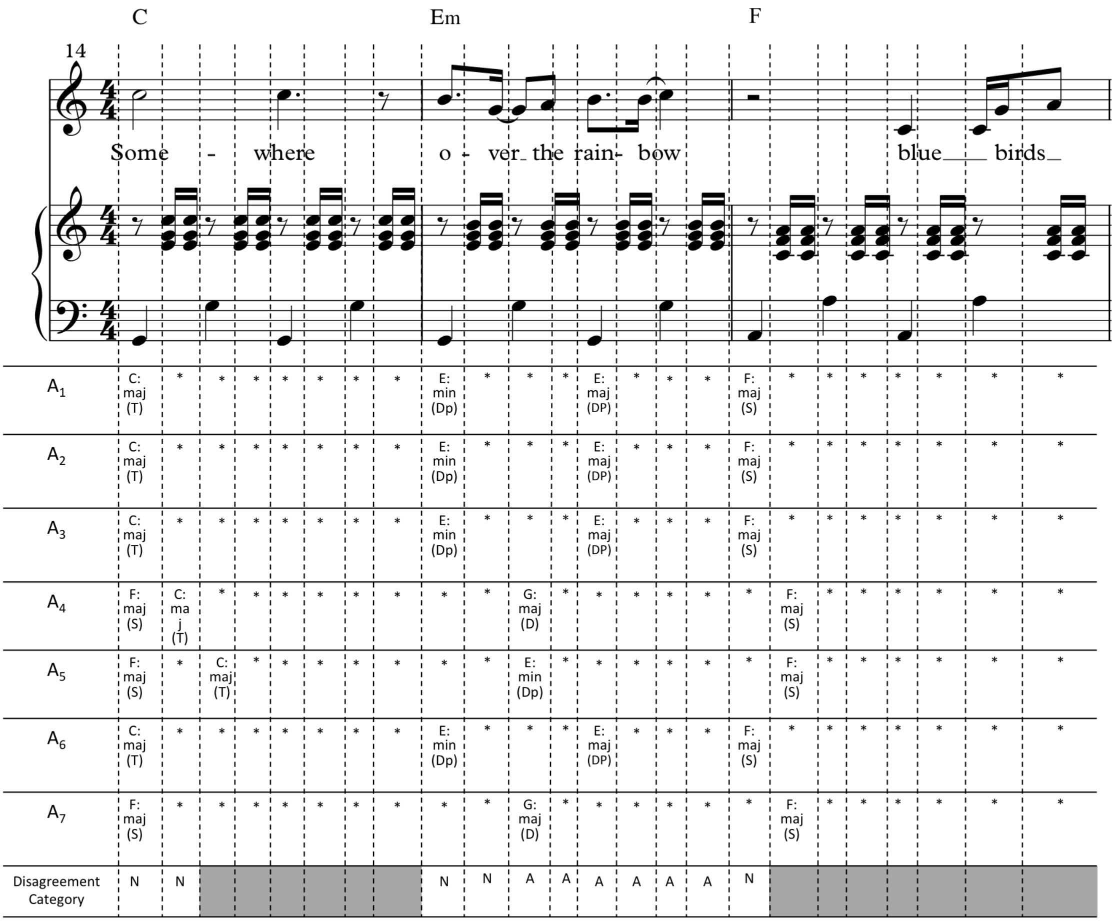


FIGURE 4.5: Bars 14 to 16 of 'Over the rainbow' by Israel 'IZ' Kamakawiwo'ole. The figure shows the musical score aligned with each of the annotators' per-beat chord labels.

contrast, 63% of the songs had no disagreement that is explainable in the category of `Partial Agreement`, including the Rolling Stones' 'Beasts of Burden'.

The `Agreement` category explains 48% of the disagreements in this dataset. The largest percentage of disagreement explained in a song was 99% for the song 'Feel it Still' by Portugal The Man, and the least was 3% for Dire Straits' 'Walk of Life'. 'Feel it Still' by Portugal The Man has such a large proportion of explainable disagreement that it is classed as an outlier (greater than two standard deviations away from the mean, with 99% of the disagreements explainable), illustrating the vast ranges of explainable disagreement. The final category, `No Agreement`, explains 50% of the disagreements in this dataset. Again, this has one outlier, 'Feel it Still' by Portugal The Man, where 1% of the disagreement is explainable by this category.

For this dataset, Riemannian theory can explain the annotator disagreement in both the categories `Partial Agreement` and `Agreement`. These two categories are combined into the category `At least Partial Agreement` to explore the percentage of explainable disagreement. This new category collectively explains 50% of the disagreements, with no outliers. Very few songs in this dataset had high results in either `At least`

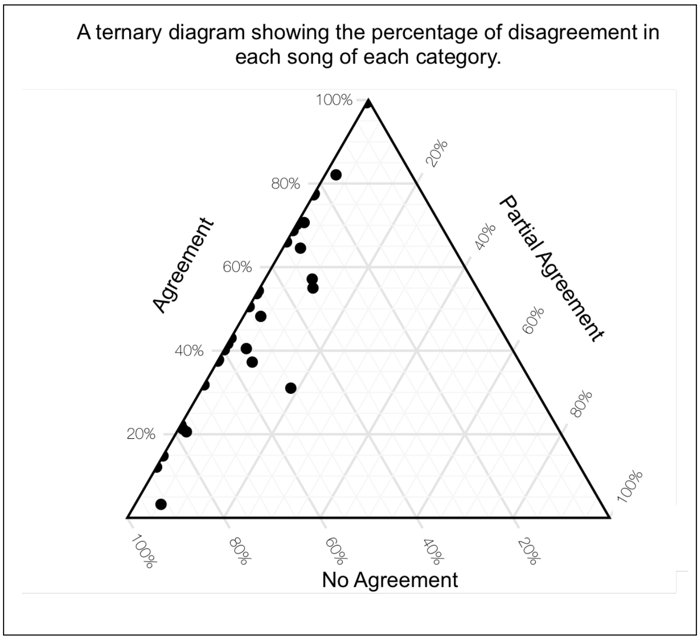


FIGURE 4.6: The percentage of disagreement, in each song, that falls into each of the three categories (`Agreement`, `Partial Agreement`, and `No Agreement`).

`Partial Agreement` or `No Agreement`, leading to this 50% average. The songs that did see extreme results included Justin Timberlake's 'Can't stop this feeling', which is 84% explainable by `At least Partial Agreement` and Tracy Chapman's 'Fast Car', of which 88% of the disagreements fall in the category `No Agreement`.

Section 4.2.1 found that elements of granularity and harmonic disagreement in the score could explain 3% of the annotator disagreement within this dataset, such as those discussed concerning the song 'Take on Me' by A-ha (Section 4.2.2.1). Points of harmonic change led to a further 15% of the harmonic disagreements — for example, the disagreement discussed for the song 'Over the rainbow' by Israel 'Iz' Kamakawiwo'ole (Section 4.2.2.4). In total, combining all the discussed methodologies, 68% of the disagreements in this dataset can be explained.

4.3 Study 2: Interviews

Following on from the first half of this chapter, which investigated the usefulness of Riemannian theory (and other musical features) for explaining harmonic disagreement between annotators, I will now investigate why and how users made harmonic annotation decisions. The concept for this additional study arose from the desire to explain the differences between the results of this chapter's analysis and Chapter 3's analysis of CASD. The biggest difference was in the amount of disagreement between the annotators that was explainable by the categories `Agreement` and `Partial Agreement`. The category `Agreement` explains the majority of the disagreement in this chapter (48%), and the category of `Partial Agreement` explains substantially less — only 2%. In contrast, `Agreement` explained 27% of the disagreement in Chapter 3, and `Partial Agreement` explained 13%. No demographic information exists for the annotators of this chapter's dataset. An interview enables us to explore possible reasons, including demographic and methodological, that the Chordify users in this chapter were less likely to disagree on the function of the harmony than the annotators in Chapter 3. Along with this, information collected on how and why annotators edit chord labels will help inform predictions as to areas of disagreement for this thesis's final study in Chapter 6.

This interview aimed to (1) understand why users decided to use the Chordify user edit function; (2) investigate the methodology used by users to annotate the harmony; and (3) explore why Chordify users were less likely to disagree on the function of the harmony than the annotators in the CASD dataset discussed in the previous chapter. The study aimed to interview as many of the 77 individuals who feature in this chapter's dataset as possible, to understand how and why they made edits to the harmonic annotations available on Chordify, along with identifying the methodology they used to make these harmonic decisions.

4.3.1 Methodology

This study uses a semi-structured interview. Unlike a structured interview, which has a specific set of questions that the interview cannot divert from, a semi-structured interview enables the interviewer to ask for further clarification from the participant, enabling the exploration of 'how' or 'why' the participant made these judgements, and to ask the participant to explain further their decision-making process.[21] An interview methodology means the investigator does not need to anticipate all possible answers to the interviews' questions. As this study is investigating individuals' processes and opinions, we cannot predict all possible answers, and an interview ensures that the answers are not limited or restricted to the investigator's preconceptions.[22]

To obtain participants, Chordify sent an email to each user that featured within my subset of their user edit data (the dataset detailed in Section 4.1.1). This email discussed the purpose of the interview, and asked the users to contact the investigator if they were interested in participating in the interview.[23] One limitation of this method was that there was no guarantee of participation from the users; also, some users no longer had active accounts and, therefore, could not be contacted. Once the participant had

21. S Soafer, "Qualitative Methods: What are they and Why use them?," *Health Services Research* 35, no. 5 (199): 1101–1118.

22. Carol A. B. Warren, "Qualitative Interviewing," in *Handbook of Interview Research*, ed. James A. Gubrium Jaber F. Holstein (California: SAGE Publications, 2001).

23. Due to GDPR requirements Chordify could not put me in contact with the users, instead the users had to communicate with me. Therefore my details were sent to them to retain their anonymity.

registered their interest, they received a consent form and information sheet for further information. Following this, we agreed a suitable interview time.

Each interview began with the participant reading a short description of the study. The interview featured mostly open-ended questions, enabling me to collect detailed data from the participants.[24] Participants were able to explain in their own words how and why they made certain decisions, and their processes, without leading questions or implicit answers.[25]

The first set of questions covered demographic matters, including age (in a closed question format), musical performance qualifications (closed question format), musical academic qualifications (closed question format), instrument (in specified response format), participation in ensembles (boolean answer format), listening habits (closed question format), and experience in harmony annotation (open response format) — these demographic questions were based on the demographic questions asked in Chapter 2. The main interview featured six questions which made up the overall structure of the interview. I followed the line of discussion of the participant, and prompted when required with further questions. The questions were:

1. How did you find the Chordify website, and what is your interest in the resources that are available on it?
2. Why do you use, or why have you used, the Chordify user edit function?
3. Can you explain how you decided which chords you wanted to change?
4. Please describe the methodology you used to determine the chord label of a segment.
5. Did you already know the pieces that you changed the chord labels of? How well did you know the pieces (e.g. had you played it before, heard them before etc.)?
6. Is there anything else you would like to add?

The interview took no longer than 45 minutes, as advised by Carol A. B. Warren (2001) and Jennifer Rowley (2012) as the maximum ideal length. The interview was kept to this length to ensure that length of process did not deter participation.[26]

4.3.2 Participants

The participants were three volunteers (4% of the participants in the dataset) over the age of 18. The participants' demographics are detailed in Table 4.2. Only one

24. Warren, "Qualitative Interviewing."
25. Jennifer Rowley, "Conducting Research Interviews," *Management Research Review* 35, nos. 3/4 (2012): 260–271.
26. Rowley, "Conducting Research Interviews"; Warren, "Qualitative Interviewing."

P	**Age**	**Musical Expertise**	**Instrument**	**Ensemble**	**Listening**	**Transcription**
P1	45–54	Self taught, "Music School" aged 20	Guitar	Yes	More than 15 hours	Arranging Film Scores
P2	45–54	Lessons aged 7–17 years	Guitar	Yes	More than 15 hours	No
P3	25–34	Father taught music & Self taught	Piano & Guitar	No - Only Jamming	10–15 hours	No

Table 4.2: The demographics of the participants for the interview study of Chordify users edit data. This table shows the participant number, age (category), musical expertise, primary instrument, ensemble experience, listening hours per week (category) and their transcription experience.

participant (P_1) had further musical training attending 'Music School' at aged 20. All three participants had some form of music education either through private lessons (P_2) or at home (P_3). All three participants played the guitar, though P_3 also played the piano. All three participants had played music with other people, though P_3 said they had only played informally in 'jamming sessions'. The participants all listened to music extensively, with over 10 hours of listening per week. Only one participant claimed any experience in annotating harmony (P_1), though, upon further discussion, their experience was in arranging film scores, not harmonic transcription by ear.

4.3.3 Results

This results section will group the results under three headings: 'Reasons for using Chordify', 'Reasons for making chord-label edits', and 'Methodology used to transcribe harmony'. These categories cover all the results of a thematic analysis, containing further subcategories within these categorical labels.

4.3.3.1 Reasons for using Chordify

The theme group 'reasons for using Chordify' aimed to explore a participant's impetus for using Chordify. The questions that featured answers under this theme heading were Questions 1, 2, 3 and 6. The themes identified here were 'Teaching', 'Partnership' and 'Autodidactism'.

a. Teaching This theme encapsulates participants' use of Chordify to teach others. Two of the participants (P_1 and P_2) taught music to beginners. Both of these participants described Chordify's usefulness for beginners to be 'able to learn chords by themselves' (P_1). However, both of these participants used Chordify differently in their teaching. Participant 1 wanted a tool to enable their students to continue their learning at home (see theme Autodidactism (c)). In contrast, Participant 2 provides online lessons for beginners to follow from home. They integrate their edited Chordify chord sequences within their lessons to enable learners to follow (something Chordify worked with them to achieve — see theme b (Partnership)). Participant 2, in particular, highlighted the usefulness of Chordify 'for people who do not know when to play the chords'. However, they discussed that it was difficult to know how many times to play a chord, and what rhythm to play, as Chordify only provides the chords on 'the first count'.

b. Partnership This theme incorporates any contractual partnerships between participants in this study and Chordify, as this could be a reason for the participants' use of the software. Participant 2 stated that they came to know of Chordify when Chordify contacted them two years ago. Chordify contacted them after the participant had already established a successful online teaching platform, leading to the participant starting to use Chordify in their business (as stated in theme a), and for their own use.

c. Autodidactism This theme encompasses the use of Chordify to complete self-paced learning — to teach yourself how to play a piece, or as part of the aim to teach yourself how to play an instrument. Participant 3 described autodidactism as their reason for finding and using Chordify:

> I remember exactly why I wanted to find Chordify ... in 2015 I was telling a [friend] about a song I really like, an R&B song from the 80s. [For] a lot of songs you can find the chord annotations online, I was searching for it for a while. I believe it [(the song)] was on YouTube. I was getting really frustrated because I wanted to learn it on guitar. This [friend] said there is probably an app you can use that would break it down and give you the bare bones of the chord structure. So, I think I probably Googled that hint blindly, and I found Chordify. It was a game-changer.

Participant 2 also stated their reason for using Chordify as autodidactism. However, they were referring to a family member who returned to playing an instrument for 'their own interest and initiative' through using Chordify. As mentioned in (a), Participant 1 used Chordify to encourage self-learning, or continued development at home for their students.

4.3.3.2 Reasons for marking chord label edits

One aim of this interview study was to understand why users make chord edits on Chordify. Answers to this theme category came from Questions 1, 2, 3, and 4. The themes identified under this category were 'correcting', 'improving', 'simplifying', and 're-aligning'.

a. Correcting All three participants used words such as 'errors', 'discrepancies', and 'faults' to describe why they might edit a song's chord labels. Participant 2 specifically identified that 'sometimes the algorithm gets it wrong'. Overall, correcting was the most popular explanation for making chord label edits. Half of the mentioned reasons for making chord label edits fell in this thematic group, and it was the only reason given by participant 1.

Participant 2 highlighted how often they felt errors arose because the 'system listens to the bass'. They gave an example of a walking bass from A–C: 'You can use a B between, or you could move to another tone; this does not mean the guitar would play that B minor chord'. It is interesting, as participant 2 highlights a disagreement between them and the algorithm on the level of granularity to annotate the harmony; whether you perceive the A as prolonged, or the harmony moving to the passing note B. In the first half of this chapter (the analysis of the subset of Chordify's user edit data), I highlighted that in this dataset disagreements on the level of granularity explained 3% of the disagreements between participants. This comment by participant 2 highlights it as a real perceptual decision that annotators had to include in their decision-making process.

b. Simplifying Participant 2 also discussed that sometimes the chords given for a piece were too difficult for beginners, and this meant they had to make changes to simplify the harmony for their students. The chords extracted by the algorithm, therefore, were correct, but just too difficult for a beginner. The participant gave the example of the chord B minor, which for an amateur guitar player would be too complicated (this is often stated as a hard chord to play because it rests on the second fret, whereas other beginner chords are open). The participant recommended substituting this B minor chord for a D major chord, as it is a simpler chord. This method of substituting the harmony relates to the methodology of the first half of this chapter and Chapter 3. The *Parallele* substitution relates D major and B minor. Similarly, this indicates a possible perceptual similarity between these chords as one can simply replace the other in a piece.

c. Re-aligning The final reason given for editing the chord labels on Chordify was to change the beat that the chord falls on; this was a particularly prominent theme in participant 3's response. Participant 2 and 3 highlighted that sometimes the timing of chords would be 'a little lagging', or 'mismatched'. One particular example Participant 3 highlighted was in the song 'I Need Your Lovin'' by Teena Marie. Interestingly, this is highlighted by participants as a reason, as Chordify removed examples where participants just moved a chord to a new segment when reducing the dataset (see Section 4.1.1). However, many songs within the dataset have disagreements that are explainable through a disagreement over when a chord change occurs (see Section 4.2.2), e.g. 'Hotel California' by the Eagles. Overall, 15% of the disagreement in this subset of Chordify's user edit data is disagreement over where a chord change occurred.

4.3.3.3 Methodology used to transcribe harmony

The final theme group answers the second aim of the study — to understand how participants made decisions when transcribing the harmony. Questions 3, 4, and 5 had responses that related to the methodology for transcribing harmony. The themes identified were 'listening', 'playing', and 'comparing'. All three participants stated they would have known the songs before they edited the harmony, though they often had not played them before and had just heard them. They usually approached a song on Chordify because they wanted to learn to play it.

a. Listening When asked to explain the methodology they used to determine the chord label of a segment, all participants stated they listened to the piece first. Some participants (P_1 and P_2) said they could hear when a chord sounded incorrect. Participant 1 stated that the changes they made were done 'by ear'.

b. Playing After listening to the recorded song, participants also stated they played the chords suggested by Chordify, and that this enabled them to hear whether they were correct or incorrect. Though Participant 2, like Participant 1, stated they made the changes by ear, they highlighted the importance of playing it on their guitar to 'check it'. This stage incorporated some aspect of listening for all participants.

c. Comparing Participants 2 and 3 went on to describe a comparative methodology either within the edits made by Chordify's algorithm or against other published chord labels. Participant 2 highlighted how they observe areas of repetition and look for differences in the repeats of the chord progressions. They stated that 'if you know the song, you know it is wrong', i.e. if a repeat of the chorus does not have the same harmony, then Participant 2 says you can assume an error in the chord labels. They gave the example of Shawn Mendes' song 'Mercy', which has the chords E minor, G major, B minor and A major repeated throughout most of the song. Therefore, Participant 2 states that if you know the song is highly repetitive, you know to check the chords are the same for each repeat.

In contrast, Participant 3 detailed how they compare the chord sequences output by Choridfy with other published chord labels. They observe the 'discrepancies' between the two and if 'the chords on another website seemed more appropriate [Participant 3] would change [the chords on Chordify]'. Participant 3 was the only participant to not re-write the harmony solely through their harmonic transcriptions by ear, but to also rely on other available resources.

4.4 Discussion and Conclusions

This chapter has presented an analysis of a subset of Chordify's user edit data using Riemann's theory of harmonic functions, as a method for explaining chord-label annotator disagreement. This chapter aimed to confirm that the methodology used to explain annotator disagreement in Chapter 3 (concerning the CASD dataset) was useful in explaining annotator harmonic disagreement in a larger dataset. Following this, an interview study explored how and why individuals make harmonic annotation decisions. Overall, Riemannian theory can explain in full 48% of the annotator harmonic disagreement, and a further 2% partially. This study confirms that Riemannian theory can explain some harmonic inter-annotator disagreement, showing a higher level of agreement between annotators at this more musically informed harmonic function level. A further 3% of this chapter's disagreement is explainable using score-based methodology (elements of granularity and harmonic disagreement). Finally, 15% of the harmonic disagreements in this dataset occurred where there is a chord change. Combining all discussed methods, 68% of the disagreements in this subset of Chordify's user edits are explainable.

Within this chapter's interview study, participants highlighted the importance of granularity and the positioning of chord changes in their decision to make chord edits. A specific example comes from Participant 2, who provided an example of annotating the harmony of a walking bass, and whether we annotate the harmony changing with the bass part; suggesting that elements of granularity may be relevant in an annotators decisions of where and what to annotate. 'Re-aligning' the chord labels was a prevalent theme in Section 4.3.3.2. Both participants 2 and 3 discussed how moving chords to the correct beat was part of their reasons for making chord label edits. Again, this emphasises the importance of disagreement over where a chord change occurs in this dataset. I found that 15% of the disagreement in this dataset fell at these points of harmonic change. In turn, this emphasises the likelihood of Chordify's beat tracking algorithm, and this study's methodology of aligning these different beat tracking algorithms, lead

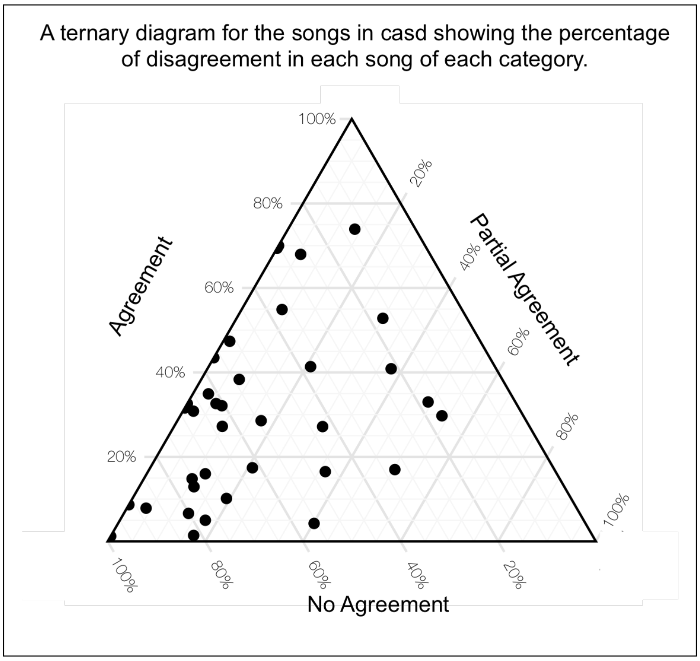


FIGURE 4.7: The percentage of disagreement in each song for CASD that falls into each of the three categories (`Agreement`, `Partial Agreement`, and `No Agreement`).

to some annotator disagreement. As this was so prominent in terms of explaining disagreement, I will look at this cause of disagreement in my thesis's last study (Chapter 6). As the score will determine the beats, I will remove the possibility that the beat tracking algorithms have caused this disagreement.

Chapter 3 used the same methodology as the first study of this chapter. When explaining the disagreements in CASD (the dataset used in Chapter 3), 40% of the disagreement could be explained using Riemannian theory, compared to the 50% explained in this chapter. Interestingly, the biggest difference between the two datasets was in the percentages explainable by the categories of `Agreement` and `Partial Agreement`. In Chapter 3, the difference between what was explainable by the two categories (27% for `Agreement` and 13% for `Partial Agreement`) was much smaller than the difference found in this chapter (48% vs. 2%).[27] Though we do not know demographic details for all the annotators within this chapter's dataset, the second study in this chapter (the interview study) introduced demographic information for those who took part in the interview. Though there were only 3 participants, and thus less than 4% of the annotators took part, the results provide an insight into the possible reasons for this difference in distribution of the `Agreement` and `Partial Agreement` categories (though we should be careful not to generalise these findings to all the participants).

This higher level of explainable disagreement in the category of `Agreement` could have arisen because the annotators are not 'expert' transcribers or annotators in this chapter's dataset. None of the interviewed participants had professional experience or training in

27. For reference, compare the two ternary diagrams, Figure 4.7 and 4.6.

transcribing harmony by ear. Two of the participants had no further musical education past 'school age'. In contrast, the annotators from Chapter 3 were all professional transcribers who had at least undergraduate training in music. Therefore, this could provide us with insight into the different ways that those formally trained, and those self-taught, hear harmony, which, in turn, may suggest that there is a tendency for those who are more specialist to have higher levels of harmonic disagreement: in other words, those with more extensive harmonic vocabularies are more likely to disagree with other annotators. I would like to highlight this as a potential avenue for further research.

Interview Participant 2 highlighted that one of the reasons that they made chord edits was to simplify the chords for their students, who are beginners. This point was of particular relevance to the methodology of Study 1 of this chapter and Chapter 3. Participant 2's answer highlighted the perceptual similarity between chords related by the *Parallele* substitution. The *Parallele* substitution was also the most frequent substitution to feature as an explanation for harmonic disagreement in Chapter 3. What is interesting in this example is that Participant 2 did not state this in relation to a key — suggesting it may not be the chords as related to the function that affect perceptions of similarity, but how the chords are related to one another. This will be further explored in Chapter 5 and Chapter 6. Chapter 5 will propose removing the need for a key in observing the similarity of chords. Finally, Chapter 6 will present an aural transcription study to explore, specifically, whether the function of a chord within a key is important for harmonic perception.

It is worth noting the limitations of this study. This dataset contains diverse popular music and annotators, however this dataset still featured a surprisingly small number of annotators per song (the largest being 11). Though this is, to date, by far the most significant number of annotators that a study of harmonic inter-annotator disagreement has observed, it is still a small sample size, and therefore raises questions on the ability to generalise our results. The first study in this chapter has no demographic details about the annotators, meaning we are unable to provide more insight into factors (e.g. primary instrument, or the proposed 'musical expertise' discussed above) that could influence chord label choice. Though I completed an interview study to provide insight into the annotators, only three annotators responded to this request, meaning we still know very little about the demographics of the participants in this dataset, and cannot generalise our demographic details to the whole dataset. Yet, importantly I have provided further qualitative evidence for the incorporation of music theory in MIR methodology.

Another limitation of this study arose from the results of the listening study. Through understanding that annotators have different reasons for annotating harmony, the assumption that the annotators disagree on the harmony based on not notating the same chord could be a simplification. The harmony that someone might annotate when choosing the most 'accurate' may, or likely, differ from the harmony they would choose to 'simplify' a piece for their students. Therefore, the annotators may also agree on the harmony of other annotators in different scenarios. Further research needs to explore and account for different motivation for editing chords in online systems.

The methodology discussed in this chapter requires a musical score to verify a Riemannian analysis (to enable us to determine any key changes within the music). Due to its recording practices, popular music is often a (subjective) transcription itself, if available at all (this has also limited the size of the dataset). Therefore, Chapter 5 will look into removing the requirement for a score, and observe substitutions (without functions) to enable this to be completed with an audio file, and to enable easier computation. Chapter 5 will use the nine pieces from Chapter 3 and the two from this chapter that had no available score. It would also be worth exploring the role that function plays in harmonic similarity; this would require the harmony in the score to be precisely the same as the audio, to make sure we can ascertain the function of the chords. My final thesis study (Chapter 6) will explore this, using a piece recorded from the score to allow for an exact replication of the audio file in a score format.

The work of Chapter 3 and this chapter has shown the importance of Hugo Riemann's theory of harmonic functions in explaining harmonic disagreement. 50% of the disagreements in this chapter could be explained at least partially by Riemann's theory. Comparing the results of this chapter with Chapter 3, this chapter showed a higher level of disagreement explained by the `Agreement` category than was explainable in Chapter 3. I suggest that this was because those with more extensive harmonic vocabularies are more likely to disagree with other annotators, suggesting that Riemannian theory explains a more substantial proportion of disagreement between amateur musicians. Through an interview study, elements of the chapter's methodology reflected participants' own methodologies for annotating harmony; again, suggesting that there is a relationship between Riemannian substitutions and harmonic disagreement. Therefore, there may be a perceptual similarity between chords of the same function. The specific results of these two chapters formulate hypotheses and predictions on where, and how annotators will disagree in Chapter 6.

Chapter 5

Adapting Hugo Riemann's theory of harmonic functions to explain inter-annotator disagreement in popular music recordings

5.1 Introduction

Music is most often stored and distributed in audio file format.[1] Popular music is rarely conceived in a notated format such as a score, whereas art music usually exists as written notation before it is performed or recorded. While published notated arrangements of pop recordings exist, these arrangements are often not written by the song's composer(s) in this medium,[2] likely having never been in a notated format during the compositional or recording process. Famously, many popular song musicians do not read sheet music, and have explicitly stated that this was not a skill expected of them in the industry. These artists include all four of The Beatles, Elvis Presley, Pete Townsend, Bob Dylan, Taylor Swift and Bob Marley.[3] Though not conceived in the format of score notation, a transcription provides a way to 'pause' the music, to enable its analysis. Currently, the musical score is generally required if one wishes to analyse a popular music piece using a music-theoretical approach, such as Riemannian theory used in Chapter 3 and 4, to identify the local and/or global key changes.

Sheet music versions of popular music exist in the form of transcriptions made to enable fans to play their favourite pieces; for example, the *Musicnotes* scores used in Chapters

1. Phillip Tagg, "Analysing Popular Music: Theory, Method and Practice," *Popular Music* 2, no. 1 (1982): 37–67.
2. Clercq and Temperley, "A corpus analysis of rock harmony."
3. Robbie Gennet, *Why is Sheet Music Still Considered Necessary for Music Education?*, Available at: https://www.huffpost.com/entry/why-is-sheet-music-still_b_7975400, December 2016.

3 and 4. These transcriptions often only constitute a 'version' of a song's most essential components, as the individual transcriber has to make judgements not only on what aspects to transcribe, but also how to arrange the transcription based on the required instrumentation. Often, the transcriber assumes that a pianist and a singer will recreate the piece; sometimes guitar tab or chord labels are included for a guitarist. However, a guitarist is likely to find the tablature or chord sequence available online (and not need the sheet music) at sites such as Chordify (see Chapter 4), and may not be able to read sheet music. These transcriptions also have to consider the difficulty level of the piece, whether an average member of the public could play it in its original format, or whether a simplification of the harmony or melodic line is required. Musical transcription is, therefore, the act of translating an existing musical work from one medium to another, with the goal of performance or (more rarely) analysis.[4]

Peter Winkler (1988) stated that transcribers only aim to have a plausible correspondence between the notated transcription and the actual recording.[5] The intention of the transcriber is, however, to be faithful to the musical content of the piece they are transcribing, while writing the transcription appropriately for the medium they are writing in.[6] Therefore, though it is necessary that the transcriber's score adequately resembles and preserves the musical content of the piece, there will be some features that are not representative of the original work.[7] David Brackett (2000) even states that 'it is, of course, impossible to present a completely "accurate" transcription',[8] which may be seen as controversial, since others would argue that it is plausible to create a score that represents the notes played within the piece. However, the score is limited in what audible features it can represent in a notated format; for example, musical features such as timbre are not represented as clearly in a score. Often, transcribers will arrange a multi-instrument piece into a piano-vocal score, which can further contribute to that transcription's divergence from the original setting. Equally, a transcriber's work may be influenced by the threshold of human perception, contextual influences, transcription errors and the complexity of the task.[9]

No specification states how far a transcriber may purposefully depart from the contents of the original piece.[10] Ferrucio Busoni even discusses the importance of incorporating a compositional element into the act of transcription.[11] Busoni is remembered as a

4. Koops, "Computational Modelling of Variance in Musical Harmony."

5. Peter Winkler, "Randy Newman's Americana," *Popular Music* 7, no. 1 (1988): 1–26.

6. Stephen Davies, "Transcription, Authenticity and Performance," *British Journal of Aesthetics* 18, no. 1 (1988): 216–227.

7. Davies, "Transcription, Authenticity and Performance"; Paul Thom, *The Musician as Interpreter* (Pennsylvania: The Pennsylvania State University Press, 2007).

8. Thom, *The Musician as Interpreter*, p. 27.

9. Stan Hawkins, "Musicological Quagmires in Popular Music," *Popular Musicology Online*, 2011, Jason Stanyek, "Forum on Transcription," *Twentieth-Century Music* 11, no. 1 (2014): 101–161; Christopher Doll, *Hearing Harmony: Toward a tonal theory for the Rock Era* (Michigan: University of Michigan Press, 2017).

10. Davies, "Transcription, Authenticity and Performance."

11. Erinn E. Knyt, "'How I Compose': Ferruccio Busoni's Views about Invention, Quotation, and the Compositional Process," *The Journal of Musicology* 27, no. 2 (2010): 224–264.

transcriber of Johann Sebastian Bach's music, and he famously utilised substantial re-composition in his transcriptions.[12] Very few transcribers agree with Busoni's approach, which allows for significant difference between the original and its transcription. Today, with such broad access to transcriptions via the medium of the World Wide Web, the public can regularly check a transcription against its original medium — there is, therefore, a higher demand for transcriptions that align accurately with their original pieces.[13] As seen in Chapter 4, in some instances, the public can now 'correct' what they perceive to be incorrect in an online transcription.

This thesis so far has used Hugo Riemann's theory of harmonic functions to explain annotator disagreement (Chapters 3 and 4). This methodology has required the use of a musical score to locate local and global key changes, to establish the functional element of Riemannian theory (as this relates to the chords' relationship to the key). As popular music is often not available in score format (for example the nine songs excluded from Chapter 3 and the two songs excluded from Chapter 4), this chapter will consider if Riemannian theory can be adapted to enable the analysis of disagreement in audio files alone. This will be achieved by removing the functional element of the analysis (tonic, dominant and subdominant) and observing only the substitutional relationship between chord disagreements (the *Leittonswechsel, Parallele* and *Variante*). The results show that the substitutional element on its own can explain at least some of the 40% harmonic disagreement in the songs discussed. As this explainable portion is comparable to the work of previous chapters, this chapter concludes that this substitution-only approach provides a promising potential for understanding inter-annotator disagreement and provides a simple computational implementation.

5.1.1 Dataset

The songs analysed in this chapter are a subset of the 11 songs from Chapters 3 and 4 that have no available score (nine from CASD the dataset used in Chapter 3, and two from the dataset used in Chapter 4). Of these songs, one, 'If I ever were a child' by Wilco (from Chapter 4's dataset), was removed from this discussion as there was no disagreement between the annotators in terms of the harmony. Three songs — 'A Trick of the Night' by Bananarama, 'Maybe Tomorrow' by Bad Finger, and 'The Look of Love' by Isaac Hayes — were also removed because the YouTube URL used to create the harmonic annotations was no longer available (the file had been removed by the content creators or by YouTube). This YouTube file was required to enable the analysis of patterns in the disagreement, as the musical form relates to the lyrical form of the song. The seven remaining songs (six from Chapter 3, and one from Chapter 4), will be discussed in this chapter.

12. Ibid.
13. Thom, *The Musician as Interpreter*.

5.1.2 Disagreement

Similarly to Chapters 3 and 4, a global disagreement analysis was used to highlight sections of overall disagreement. All-in-all, there is a 44% disagreement between the annotators (there were four annotators for six of the songs, and three annotators for one song). In Chapter 3, a disagreement of 34% was found, and in Chapter 4 there was a disagreement of 18%; but the global disagreement found in this chapter is the highest level detailed thus far in this thesis. This higher level of disagreement may be due to the complexity of the songs, which may also explain why they have not been transcribed for fans to play.

5.2 Methodology

The methodology used in this chapter is similar to that of Chapters 3 and 4. However, since this chapter does not make use of musical scores, no harmonic annotations are required for the purposes of alignment. The six songs from Chapter 3 also did not require annotation alignment, since the conditions used to create the dataset lead to the annotations already being aligned (see Chapter 3, Section 3.1.2 for more details on the creation of this dataset).[14] The song from Chapter 4 required further alignment, since the annotators may have annotated the harmony using different beat-tracking algorithms (see Chapter 4 Section 4.2.1 for a detailed discussion of this process).

This chapter's methodology focused on categorising the annotators chord labels. First, each set of chord labels (a segment) were observed to see if there were any disagreements between the annotators. Those segments that featured harmonic disagreement were then categorised based on whether each unique pair of chords in the disagreement are related by a single substitution. The unique pairs are established by first observing the unique chords in the disagreement: for example, in the collection of chords' C major, C major, G major, and E minor, the unique chords are C major, G major, and E minor. Following this, all the unique combinations of chords (for this example, these pairs are: C major with G major, C major and E minor, and G major and E minor). Then, each pair of chords is examined to see if they are related by a single substitution. Disagreement is then defined using the following categories, based on the amount of disagreement between the unique pairs that are related by a single Riemannian substitution: `Related`, `Not Related`, and `Partially Related`.

The category `Related` is used in a similar manner to the `Agreement` category from Chapters 3 and 4. This category shows that all unique pairs of chords were related by only a single substitution (*Parallele*, *Leitonswechsel*, or *Variante*).[15] For this dataset,

14. Hendrik Vincent Koops et al., *Harmonic Subjectivity in Popular Music*, technical report UU-CS-2017-018 (Department of Information and Computing Sciences, Utrecht University, 2017).
15. see Chapter 3 section 3.2.1 for a revision on Riemannian substitutions

the disagreement is usually between two chords, for example C major (perceived by A_1 and A_2) and A minor (perceived by A_3 and A_4). These chords are related by the *Parallele* substitution and are, therefore, categorised as `Related`.

The category `Not Related` explains segments where none of the unique chords are related by a single substitution: for example, C major and B♭ major are not related to each other. This category often includes three or four chords in the disagreement.

It is more complex to categorise the disagreement where there are more than two chords involved (i.e. three or four chords, as the dataset featured no more than four annotators per song). To categorise these disagreements, I adapted the category `Partial Agreement` from Chapters 3 and 4, and renamed it `Partially Related` and alongside this I have created a new category called `Some Relation`. `Partially Related` explains situations where over half of the unique chord labels are related by a single substitution. For example, to calculate the percentage of agreement between the chords C major, A minor, and E minor, we compare each unique combination of distinct chord labels — C major and A minor, C major and E minor and, E minor and A minor. The number of these distinct pairs that are related by a single substitution is then counted, and divided by the total number of distinct pairs. The number of distinct pairs, that can be explained by a single substitution in this instance, is two: the pair C major and A minor by the *Parallele* substitution and the pair C major and E minor by the *Leitonswechsel* substitution. Therefore, there is a percentage agreement of 67% between the annotators (we can explain two out of three of the distinct chord pairings). As we can explain greater than 50% of the annotator disagreement, we categorise this segment as `Partially Related`.

If the percentage of distinct chord label pairs that are explainable by a single substitution is less than 50%, the segment is categorised as `Some Relation`. For instance, the chords C major, A minor and B major, have only one distinct pair (C major and A minor) that is explainable by a single substitution. Therefore, we can explain 33% of the distinct pairs using Riemannian substitutions, and the segment can be categorised as `Some Relation`.

5.3 Results: Analyses

This section will discuss all seven pieces and how this adapted methodology can explain some of the harmonic disagreement present within each piece. I will discuss all the pieces as the methodology for each extract varied depending on the resources available. Other methods of explaining disagreement determined in Chapters 3 and 4, such as the return of disagreement with the reoccurence of a prominent feature of the piece, will also be highlighted in each song. I use both the lyrics and the audio file to highlight where the disagreement occurs, and whether this disagreement links to a specific section of the piece's form. Finally, I note disagreements that occur at points of harmonic change, as this was a particularly prominent feature of the dataset used in Chapter 4 (explaining 15% of the annotator disagreement).

5.3.1 'Dial my Heart' by The Boys

The first song, 'Dial my Heart' by The Boys, is from the CASD dataset, used in Chapter 3. This song has 57% disagreement between the four annotators, of which 9% is explainable by the `Related` category, and 47% is explainable by the `Partially Related` category. The `Some Relation` category does not explain any of the remaining disagreements. Overall, 56% of the disagreement can be explained at least partially in this song using Riemannian substitutions.

Within the `Related` category, 83% of the explainable disagreements used the *Leittonwechsel* substitution, and 17% used the *Parallele* substitution. The *Variante* substitution did not explain any of the annotator disagreement in 'Dial my Heart'.

Chapters 3 and 4 found that when disagreement occurred in highly repeated sections of a song, the disagreement also occurred with that sections' return. We can see this is the case in 'Dial my heart', as 34% of the song's disagreement occurs in the chorus sections. A chorus, by its definition, is repeated throughout a song after each verse. The disagreement present in the chorus of 'Dial my Heart' returns with each reprise of the chorus. Table 5.1 shows the annotators disagreement, aligned with the chorus lyrics, along with the segment numbers and time of the disagreement for each reprise. The disagreement between the chords E minor9, B minor7 and G major, is first present on the lyrics 'Got to get a message just to let her know', and then repeated on the second line of the chorus section 'You can reach me, baby, by the nearest payphone', and finally this disagreement is heard to the title lyrics 'my heart' in both iterations during the chorus (see Table 5.1). Between these three chords, 67% of the disagreement is explainable, as a single Riemannian substitution does not relate the unique chord pair of E minor and B minor. Therefore, we categorise this disagreement as `Partially Related`.

The other prominent disagreements in this song are between the chords G major, G major7, and B minor, on the lyrics 'to' and between the chords G major7, D major/3, and B minor on the lyrics 'dial'. G major and G major7 are equivalent, and both are related to B minor by the *Leitonswechsel* substitution, categorising the disagreement on the lyrics 'to' as `Related`. The disagreement on 'dial', between G major7, D major/3, and B minor is categorised as `Partially Related`, as G major is related to B minor by the *Leitonswechsel* substitution, and D major is related to B minor by the *Parllalele* substitution. However, D major is not related to G major. Therefore, for the 'dial' segments, we can explain 67% of the disagreement in this song using Riemannian substitutions. The disagreement in the chorus alternates with segments of agreement on the chord A major; this features on the lyrics 'I call my baby on the phone', 'That if she ever feels alone', 'And if by chance you're not alone', and 'I gotta cross my fingers and pray you'll know too'.

Table 5.1: The lyrics of the chorus of 'Dial my Heart' by The Boys, aligned with the segments, time and the four annotators' chord labels.

Beginning of Table						
Lyrics	**Seg no.**	**Time**	**A1**	**A2**	**A3**	**A4**
I call my baby on the phone.	101-104 197–200 293–296 325–328 389–392	00:58-01:00 01:55– 01:57 02:53–03:55 03:12–03:14 03:50–03:52	A:maj/2	A:maj	A:maj/2	A:maj
Got to get a message just to let her know,	105–108 201–204 297–300 329–332 393–396	01:00–01:02 01:58–02:00 02:55–02:57 03:14–03:16 03:52–03:54	E:min9	B:min7	E:min9	G:maj
That if she ever feels alone,	109–112 205–208 301–304 333–336 397–400	01:03–01:05 02:00–02:02 02:57–02:59 03:17–03:19 03:55–03:57	A:maj/2	A:maj	A:maj/2	A:maj
To	113 209 305 337 401	01:05 02:03 03:00 03:19 03:57	G:maj7	G:maj	G:maj7	B:min
dial	114 210 306 338 402	01:06 02:03 03:00 03:20 03:58	G:maj7	D:maj/3	G:maj7	B:min
my heart.	115–116 211–212 307–308 339–340	01:06–01:07 02:04–02:05 03:01–03:02 03:20–03:21 03:58–03:59	G:maj7	E:min	E:min9	B:min

Continuation of Table 5.1						
Lyrics	**Seg no.**	**Time**	**A1**	**A2**	**A3**	**A4**
And if by change you're not at home,	117–120 213–216 309–312 341–344 405–408	01:08–01:09 02:05–02:07 03:02–03:04 03:21–03:23 04:00–04:02	A:maj/2	A:maj	A:maj/2	A:maj
You can reach me, baby, by the nearest payphone.	121–124 217–220 313–316 345–348 409–412	01:10–01:12 02:07–02:09 03:05–03:06 03:24–03:26 04:02–04:04	E:min9	B:min7	E:min9	G:maj
I gotta cross my fingers and pray you'll know	125–128 221–224 317–320 349–352 413–416	01:12–01:14 02:10–02:12 03:07–03:09 03:26–03:28 04:04–04:06	A:maj/2	A:maj	A:maj/2	A:maj
To	129 225 321 353 417	01:15 02:12 03:09 03:29 04:07	G:maj7	G:maj	G:maj7	B:min
dial	130 226 322 354 418	01:15 02:13 03:10 03:29 04:07	G:maj7	D:maj/3	G:maj7	B:min
my heart.	131–132 227–228 323–324 355–356 419–420	01:16–01:17 02:13–02:14 03:11–03:12 03:30–03:31 04:08–04:09	G:maj7	E:min	E:min9	B:min
End of Table						

5.3.2 'Where the Streets have no Name (Can't Take my Eyes off You)' by Pet Shop Boys

The second song also comes from the CASD dataset (Chapter 3): Pet Shop Boy's mashup[16] of 'Where the Streets have no Name' by U2 and 'I can't Take my Eyes off

16. A mashup, in music, is a mixture or fusion of elements of two or more pieces of music.

You' by Frankie Vallie, which was famously covered by Frank Sinatra with the alternative title 'I Love You, Baby'. This song features 69% disagreement between the four annotators on the chord labels. Of this disagreement, 20% is categorised as `Related`, 3% as `Partially Related`, 3% as `Some Relation` and 69% as `Not Related`. In total, 26% of the disagreement features at least some disagreement that can be explained by Riemannian substitutions. Of the remaining disagreement, 3% occurs at a point of harmonic change, meaning the annotators agree on the harmony before and after the segment, but they disagree on the exact point that the chord changes from one chord to the next (this is a particularly prominent explanation of disagreement in Chapter 4).

Of the fully explained disagreement (`Related`), 57% uses the *Parallele* substitution, mostly between the chords G minor and B♭ major, and 20% uses the *Variante* substitution mostly for explaining the disagreement between C minor and C major. Finally, 13% of the explainable disagreement uses the *Leitonswechsel* substitution.

The 3% disagreement explained in the `Some Relation` category is between the chords G♭ minor, D♭ major, E♭ minor and B♭ minor — i.e. all four annotators perceive different chords. Out of the six distinct pairings of these four chords, we can explain two pairs using Riemannian substitutions — G♭ minor and D♭ major (*Leitonsewchsel*), and D♭ major and B♭ minor (*Parallele*). In total, 33% of the disagreement between annotators can be explained, categorising the disagreement as `Some Agreement`.

The explained disagreement between the chords G♭ major and E♭ minor (the *Parallele* substitution), is repeated throughout two sections in the second half of the song (at 01:56 and 03:27 lasting approximately 10 seconds each).[17] The disagreement is on the famous brass instrumental interlude, from Frankie Valli's 'I can't take my eyes off you', which introduces the chorus 'I love you, baby'. In the Pet Shop Boy's piece, this instrumental interlude transitions from 'Where the streets have no name', into 'Can't take my eyes off you'.

Though a score of the Pet Shop Boy's mashup does not exist, many transcriptions of the brass part from Frankie Valli's original piece 'Can't Take my Eyes off You' are available, such as, for example, Figure 5.1 (this example is in a different key to that of the recorded track that the participants heard). The melody line of the original moves in chromatic steps, which distorts the harmony. The disagreement between the annotators in the Pet Shop Boys' song aligns with the bars featuring repeated chromatic steps (bars 1 and 3, and the first half of bar 2). This riff causes 4% of the overall unexplained disagreement between the annotators (which is similar to the disagreement in Rick James' 'Super Freak' in Chapter 4, which explained 28% of the songs disagreement).

17. YouTube URL: `https://www.youtube.com/watch?v=Jt2j79pca7c`

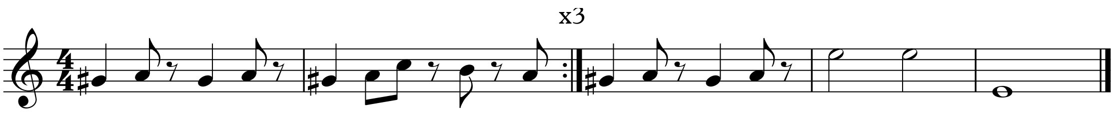


FIGURE 5.1: The brass part of Franki Valli's 'Can't Take my Eyes off You', showing the bars of semitone movement.

5.3.3 'Someone' by the Rembrandts

The third piece, 'Someone' by the Rembrandts, again comes from the CASD dataset (from Chapter 3). 'Someone' features 42% disagreement between the four annotators. We cannot explain 90% of the disagreements using Riemannian substitutions (`Not Related`). None of the disagreements are partially or fully explainable according to chords related by Riemannain substitutions (`Partially Related` and `Related`). Only the remaining 10% of annotator disagreements are explainable according to chords related by Riemannian substitution (`Some Relation`). 8% of the unexplainable disagreements can be explained by their occurrence at points of harmonic change.

Prominently, 16% of the disagreements in this song occur in the chorus, specifically, at the beginning of the lines starting with the title lyric 'Someone...'. The chorus is repeated five times with the following lyrics (I have numbered the lines for ease in the following discussion):

1. Someone to hold me, the way that you do.
2. Someone who needs me, the way that I need you.
3. Someone to show me, a way that is true.
4. Someone to love me, the way that I love you.

Disagreement never falls on the fourth line of the chorus, and none of the disagreements are repeated in all five repetitions of the chorus. The most consistent disagreement is on the second line, 'Someone who needs me', where four out of five of the choruses include this disagreement (all but the first chorus). The second most frequent line to feature harmonic disagreement is line 3, 'Someone to show me', which features disagreement in three out of the five of the repetitions. Finally, the first line features disagreement for two of the repetitions.

The disagreements all feature the chords C♯ major/3,[18] and D♯ major/2. The third chord in the disagreement alternates between F♯ major, E major, and B major — A_4

18. This form of popular music notation details the bass note of a chord, so a chord labelled as /3 means that the third note is in the bass of the music.

was the annotator to change between the three chords. There is no consistent pattern as to which chord A_4 chooses, though most frequently they chose F♯ major. The different sets of three chords (in the disagreement) are unrelated in terms of Riemannian substitutions. There are also no common tones between C♯ major and D♯ major (the two chords that feature in every disagreement). The repeated disagreement found within the chorus sections accounts for 16% of the disagreements that could not be explained by Riemannian substitutions (`Not Related`). This disagreement follows the same patterns found in Chapters 3 and 4, which show that disagreement that occurs in a repeated section, such as a chorus, is likely to continue throughout the song.

5.3.4 'Too Weak to Fight' by Clarence Carter

The fourth extract, 'Too Weak to Fight' by Clarence Carter, is again from the CASD dataset, from Chapter 3. There is a 42% disagreement between the four annotators, of which, we can fully explain 84% using Riemannian substitutions (the category `Related`). Of the 84% disagreement that features chords related by Riemannian substitutions (`Related`), 98% is explainable using the *Paralelle* substitution: for example, the disagreement between the chords D♭ major and B♭ minor/3, and the disagreement between the chords G♭ major and E♭ minor/3. The *Parallele* substitution was also the most popular substitution to explain disagreement in Chapter 3. Throughout the song, 'Too Weak to Fight', it is A_3 who disagrees with the other annotators in terms of the harmony.

The majority of this song's disagreement occurs in the verse, which repeats four times. The song's structure is: verse, verse, chorus, verse, verse, chorus, bridge, chorus. The first disagreement occurs on the opening words of the song 'There is something' at 00:09.[19] In the first two verses, the disagreements fall mostly at the beginning and/or end of the line. Disagreement is indicated by the lyrics in bold, below:

There is something baby about you,
That's really attracting me, **yeah**.
And your sweet love **darling**,
Really got a hold on me yeah.

[Instrumental]

I've got a little taste of your love,
And now I'm hooked on **you, yeah**.
And I keep **falling**,
Falling but what can I do?

19. Hogames, 'Clarence Carter - Too Weak to Fight (Original Version)', YouTube, 9 December 2011, https://www.youtube.com/watch?v=WVDSKaua0_U [accessed 25th January 2020].

Lyrics	Seg no.	Time	A1	A2	A3	A4
There is something	17–18	00:10–00:11	C♯:maj	C♯:maj	B♭:min/♭3	C♯:maj
baby	19–20	00:11–00:12	C♯:maj	C♯:maj	C♯:maj	C♯:maj
about you,	21–24	00:12–00:14	F♯:maj	F♯:maj	F♯:maj	F♯:maj
That's really attracting	25–26	00:14–00:15	C♯:maj	C♯:maj	B♭:min/♭3	C♯:maj
me,	27–28	00:15–00:16	C♯:maj	C♯:maj	C♯:maj	C♯:maj
yeah.	29–30	00:17–00:18	F♯:maj	F♯:maj	E♭:min/♭3	F♯:maj
And your	33–34	00:19–00:20	C♯:maj	C♯:maj	B♭:min/♭3	C♯:maj
sweet love	35–36	00:20–00:21	C♯:maj	C♯:maj	C♯:maj	C♯:maj
darling,	37–38	00:21–00:22	F♯:maj	F♯:maj	E♭:min/♭3	F♯:maj
Really got a hold on	39–40	00:22–00:24	F♯:maj	F♯:maj	F♯:maj	F♯:maj
me, yeah.	41–44	00:24–00:25	B♭:min7	B♭:min7	B♭:min7	B♭:min7
Instrumental	45–46	00:26–00:27	F♯:maj	F♯:maj	E♭:min/♭3	F♯:maj
I've got a little	49–50	00:28–00:29	C♯:maj	C♯:maj	B♭:min/♭3	C♯:maj
taste of your love,	51–52	00:29–00:30	C♯:maj	C♯:maj	C♯:maj	C♯:maj
And now	53–54	00:31–00:32	F♯:maj	F♯:maj	E♭:min/♭3	F♯:maj
I'm hooked on	55–56	00:32–00:33	F♯:maj	F♯:maj	F♯:maj	F♯:maj
you,	57–58	00:33–00:34	C♯:maj	C♯:maj	B♭:min/♭3	C♯:maj
yeah.	61–62	00:35–00:36	F♯:maj	F♯:maj	E♭:min/♭3	F♯:maj
And I keep	63–64	00:36–00:37	F♯:maj	F♯:maj	F♯:maj	F♯:maj
falling.	65–66	00:38–00:39	C♯:maj	C♯:maj	B♭:min/♭3	C♯:maj
Falling	67–68	00:39–00:40	C♯:maj	C♯:maj	C♯:maj	C♯:maj
but what can I do?	69–70	00:40–00:42	G♯:maj	G♯:maj	G♯:maj	G♯:maj

TABLE 5.2: The lyrics of the first two verses of 'Too Weak to Fight' by Clarence Carter aligned with the segments, times of the segments and the chords transcribed by the annotators. Disagreement that cannot be explained is highlighted in bold.

Table 5.2 shows the lyrics and chords as perceived by each annotator in the first two verses. Their annotations alternate between agreement and disagreement every couple of seconds (agreement on C♯ major and F♯ minor, for example, on the lyrics 'baby', and 'about you'). Between the two verses the disagreement is similar. It occurs at the beginning of the first ('There is something') and the second lines ('That's really attracting'), and the end of the second ('yeah'), and third lines ('darling'). Disagreement is also present in the instrumental part that connects the two verses. The disagreement continues similarly for the remaining verses, falling at the beginning and/or end of the line. Disagreement that occurs at the beginning or end of a line is not a feature that has been prominent in the previous two chapters (Chapters 3 and 4), but this feature explains 43% of the disagreement in this chapter, and therefore shows an important role in explaining annotator disagreement. It is worth remembering that only A_3 disagrees with the other annotators in terms of the harmony, and therefore it is their disagreement that is most explainable using this method.

5.3.5 'Baby I'm burnin'' by Dolly Parton

The fifth song, 'Baby I'm burnin'' by Dolly Parton, is again from the CASD dataset. The song features 63% disagreement between the four annotators. None of this disagreement is explainable using Riemannian substitutions, meaning all the disagreement falls within the category `Not Related`. Of the 63% harmonic disagreement, 10% happens at points of harmonic change, meaning agreement between the annotators exists on the chord before and after the segment, but the annotators disagree on which segments the harmony changes.

Of the disagreement, 45% is between the chords B major and E major, all within the first 1 minute 49 seconds of the 2 minutes 41-second piece.[20] The disagreement changes after 1 minute 40 to between the chords C major and G major. Leading up to 1 minute 40 seconds (a repeat of the chorus), Dolly Parton sings the lyrics 'Baby I'm burnin'' four times, raising the pitch with each repetition; this leads into a key change, and then a return to the chorus in this new key (if we had a score we would expect a key change at this point). The disagreement is repeated in the choruses throughout the remainder of the song, but interjected with annotator agreement on the C major chord at the words 'Baby I'm burnin''.

Table 5.3 shows the disagreement present in the first chorus after the key change (starting at 01:39). This disagreement is repeated for the choruses starting at 1:54, 2:08 and 2:22. The chorus disagreement occurs in every other row (in Table 5.3), and in the second half of the segment, never on the title lyrics 'Baby I'm burnin''. Specifically, this disagreement occurs at the beginning and end of the second phrase; for example on the words 'out',

20. littlesparrow185, 'Dolly Parton 06 Baby I'm Burnin'', YouTube, 11 July 2012, https://www.youtube.com/watch?v=nu6VbUAhs1M [accessed 25th January 2020].

Lyrics	Seg no.	Time	A1	A2	A3	A4
Baby I'm burnin'	223–224	01:39–01:40	C:maj	C:maj	C:maj	C:maj
out	225–226	01:40	G:maj	G:maj	G:maj	C:maj
of	227–228	01:41	C:maj	C:maj	C:maj	G:maj
control,	229–230	01:42	F:maj	F:maj	F:maj	G:maj
Baby I'm burnin'	231–232	01:43–01:44	C:maj	C:maj	C:maj	C:maj
body	233–234	01:45–01:47	G:maj	G:maj	G:maj	C:maj
and	235–236	01:45–01:47	C:maj	C:maj	C:maj	G:maj
soul.	237–238	01:45–01:47	F:maj	F:maj	F:maj	G:maj
Hot as a Pistol	239–240	01:47–01:48	C:maj	C:maj	C:maj	C:maj
that	241–242	01:48–01:50	G:maj	G:maj	G:maj	C:maj
flamin'	243–244	01:48–01:50	C:maj	C:maj	C:maj	G:maj
desire.	245–246	01:48–01:50	F:maj	F:maj	F:maj	G:maj
Baby I'm burnin	247–248	01:50–01:51	C:maj	C:maj	C:maj	C:maj
you	249–250	01:51–01:54	G:maj	G:maj	G:maj	C:maj
got me	251–252	01:51–01:52	C:maj	C:maj	C:maj	G:maj
on fire.	253–254	01:51–01:52	F:maj	F:maj	F:maj	G:maj

TABLE 5.3: The lyrics of the chorus of 'Baby I'm Burnin'' by Dolly Parton aligned with the segments, time sections, and the chord labels given by each annotator. Disagreement that cannot be explained is emphasised using bold text.

and 'control' in the first line. Again this shows how disagreement occurs at the beginning or end of phrases in examples in this chapter. Dolly Parton's 'Baby I'm burnin'' is also a strong example of repeated disagreement due to a reoccurring section; after the 1:40 key change, the disagreement within the chorus continues throughout the remainder of the song.

5.3.6 'For Ol' Times Sake' by Elvis Presley

The next song, Elvis Presley's 'For Ol' Times Sake', (from the CASD dataset used in Chapter 3) features 29% disagreement between the four annotators, of which 34% is fully explainable as chords related by Riemannian substitutions (`Related`). A further 3% of the harmonic disagreement featured some disagreement that is explainable using Riemannian substitution (`Some Relation`).

Of the 34% that can be fully explained as chords related by Riemannian substitutions (`Related`), 84% used the *Leitonswechsel* substitution, 14% used the *Parallele* substitution, and 2% used the *Variante* substitution.

In terms of the unexplained disagreement (`Not Related`), 33% features the chords E♭ major and B♭ major, and 22% features the chords F major and B♭ major. This song's disagreement mostly occurs when the music is moving into, or out of, instrumental sections. For example, the disagreement between B♭ major (A_1, A_2, A_3) and E♭ major

(A_4) is at 00:34 as the bridge begins,[21] to the lyrics "Cause it would be a shame if you should leave'. The piece ends with repeated disagreement on the title lyrics 'let me hold you for ol' times sake', from 02:51. The disagreement falls on all the words except 'you' and 'sake': It is between F major (A_1, A_3) and B♭ major (A_2 and A_4) to the lyrics 'let me hold', and F major (A_4) and B♭ major (A_1, A_2, A_3) to the lyrics 'Ol' times' at 02:51. Then again at 03:15.

5.3.7 'Laura no està', by Nek

'Laura no està', by Nek, is from the dataset used in Chapter 4 (the only song in this chapter from Chapter 4). There is only an 8% disagreement between the three annotators, of which Riemannian substitutions can explain 72% (`Related`), reflecting the higher level of disagreement explainable in the category `Agreement` in Chapter 4's dataset than in Chapter 3's dataset. The remaining 28% is explainable using Riemannian substitutions (`Not Related`). Interestingly, we can explain a further 25% of the disagreement as occurring at points of harmonic change — In total, 97% of the disagreements are explainable.

Of the 72% harmonic disagreement that can be explained as related by Riemannain substitutions, 65% are explained by the *Parallele* substitution, 30% by the *Leitonswechsel* substitution, and 4% by the *Variante* substitution.

The harmonic disagreement in 'Laura no està' does not occur at structural points in the song, or follow any strict pattern. The disagreement most often occurs at points of harmonic change; this is particularly interesting, because the song originates from the dataset used in Chapter 4. In Chapter 4, 15% of the harmonic disagreements occurred at points of harmonic change. 'Laura no està' continues this pattern, and 25% of the disagreement have occurs at points of harmonic change.

5.4 Discussions and Conclusions

This chapter has presented an analysis of the seven songs from the CASD dataset (used in Chapter 3) and the Chordify user edit dataset (from Chapter 4) that had no available transcription. This chapter aimed to explore whether an adaption of Riemannian theory that removes the functional element of the theory is useful in explaining harmonic disagreement on a dataset with no available scores. Overall, Riemannian substitutions explain at least some of 40% of the harmonic disagreement; this is similar to the 40% that is explainable, at least partially, using Riemannian substitutions in Chapter 3 (relating to the CASD dataset). As 86% of the songs in this chapter are from the CASD

21. annieparadowska, 'Elvis Presley For Ol' Times Sake', YouTube, 28 July 2009, https://www.youtube.com/watch?v=6ZINvp6SEw4 [accessed 25th January 2020]

dataset, it is not surprising that, on average, we found the same amount of explainable disagreement. This chapter has shown that Riemannian substitutions can provide an explanation for some harmonic inter-annotator disagreement, showing a similarly high level of agreement between annotators at this more 'musically informed' level. A further 5% of the harmonic disagreement in this dataset occurs at points of harmonic change. This disagreement arises at places where the harmony changes and does not concern chords labelling. Combining both methods, in total, I can explain 45% of the disagreement in these seven songs.

'Laura no està' was from the Chordify user edit dataset from Chapter 4. This chapter had an average of 50% explainable harmonic disagreement using Riemannian theory (category `Agreement`). In contrast 'Laura no està' had 72% explainable disagreement using harmonic substitutions, which is higher than the average for Chapter 4. Interestingly, this song only had disagreement that was explainable in the `Related` category, with no `Partially Related` or `Some Relation` segments; a common pattern of the dataset in Chapter 4. Chapter 4 had a higher level of explainable disagreement than was apparent in Chapter 3, suggesting a possible relationship between increased musical expertise and higher levels of disagreement. The results of 'Laura no està' suggest that this higher weighting of the category `Agreement` or `Related` is still present in the analysis of disagreement when a score is not available — in turn, this suggests that a similar amount of disagreement is explainable when no score is available.

Interestingly, it is not just 'Laura no està' that features higher levels of the `Related` category than the categories `Partially Related` or `Some Relation`: in this chapter disagreements can be explained 31% of the time in the `Related` category, 7% of the time in the `Partially Related` category, and 2% in the `Some Relation` category. This much higher occurrence of the `Related` category aligns more closely with Chapter 4 than CASD's lower difference between the `Agreement` and `Partial Agreement` categories.

Within the `Related` category, an average of 4% of the disagreement was explained by the *Variante* substitution, 31% by the *Parallele* substitution and 35% by the *Leitonswechsel* substitution. Chapter 3 saw the *Parallele* substitution as the most frequent substitution to feature as an explanation of harmonic disagreement. This chapter shows the *Leitonswechsel* as the most frequent substitution, and the *Parallele* as another frequently used substitution. This finding aligns with the work of Carol Krumhansl (1988) who shows the relationship between substitution-related chords is likely due to the chords having two common tones, requiring only a single pitch shift.[22] However, as discussed in Chapter 3, the idea that this means we perceptually hear a similarity between the chords related by one substitution was previously speculative; but my findings suggest that this idea holds true in practice.

22. Krumhansl, "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations."

In a similar way to the previous two chapters (Chapters 3, and 4), disagreement that occurs in sections of the song that are repeated (for example the verse or chorus), or within an instrumental riff, is generally also repeated. The songs 'Dial my Heart' by The Boys, and 'Someone' by the Rembrandts both feature disagreement repeated with each chorus. The disagreement in the chorus explains 34% of the disagreements in 'Dial my heart' and 16% of the disagreements in 'Someone'. Disagreement that occurs on a musical riff, and repeats with the riff, was prominent in the song 'Where the streets have no name (Can't take my eyes off you)' by the Pet Shop Boys. Two distinct segments of disagreement are present during the famous bass riff from 'Can't take my eyes off you'. Disagreement arising in sections that repeat, and then returning with the repetition of a section or riff, is a prominent type of disagreement in the songs in all three chapters (Chapters 3, 4 and this chapter). Therefore, this cause for disagreement acts as a predictor of possible disagreement segments in this thesis's final study, detailed in Chapter 6.

In these seven songs, 43% of the harmonic disagreement occurred at the beginning or end of a phrase (such as in 'Too weak to fight', 'Someone', and 'Baby I'm burnin''). This cause for disagreement is not something highlighted in the previous two chapters. It will be interesting to see if this disagreement pattern occurs in other datasets for example, as a predictor for disagreement in this thesis's final study (Chapter 6).

This chapter has built on the work of Chapters 3 and 4 in demonstrating the importance of Hugo Riemann's theory of harmonic functions in explaining some harmonic disagreements. The substitution element alone can explain harmonic disagreement, suggesting that there is a direct relationship between Riemann's concept of harmonic substitutions and annotator disagreement, and that there may be a perceptual similarity between chords related by a substitution. Additionally, this chapter also shows that disagreement over where a chord change takes place and disagreement that reoccurs with a repeated musical element, were also significant for explaining annotator disagreement. Due to the success of this chapter's methodology, Chapter 6 will explore if the functional element of Riemann's theory explains any more of the disagreements than the substitution-only approach used here. I will do so by guaranteeing the local and global keys of the piece through the creation of my own audio recording of a song. The prominent features of the disagreements discussed in these three chapters (Chapters 3, 4 and 5) will be used to predict elements of disagreement in the song used in this thesis's final chapter (6). This chapter's conclusion about the usefulness of Riemannian substitutions in explaining harmonic disagreement in audio files provides an easy computational implementation, which I discuss in Chapter 7. Overall, this chapter has shown the promising potential for the understanding of inter-annotator harmonic disagreement and its implementation in MIR tasks such as ACE.

Part III

Chapter 6

The impact of score accuracy on Riemannian theory's ability to explain annotator harmonic disagreement.

6.1 Introduction

Music theory and analysis often involves performing 'close readings' of pieces of music using the score.[1] Hugo Riemann's theory of harmonic functions is no different: the theory requires a score to confirm a piece's local and global key changes, and in turn, to work out the harmonic functions. The scores used so far in this thesis have themselves been transcriptions, created after the recording, meaning they could have been affected by the subjective nature of transcription (Chapters 3 and 4). Therefore, we cannot guarantee that the scores used in these chapters accurately represent the audio. In Chapter 5, therefore, I compared a substitution-only approach with a harmonic function approach using potentially unreliable scores. In this chapter, I remove the unreliability of the score by recording my own version of a piece of music, which was conceived first in a score-based format. I aim to confirm whether harmonic function (the aspect of Riemannian theory that requires the musical score) affects the amount of explainable disagreement when using Riemannian theory, or whether we can rely on an audio-only/substitution-only approach as detailed in Chapter 5.

Additionally, this chapter also aims to confirm the results of the previous three chapters (Chapters 3, 4 and 5) by using the explainable disagreement to predict where, and what types of disagreement will arise. A review of the previous chapters' findings (Section

1. Bent and Pople, "Analysis."

6.1.1) will highlight five causes of disagreement and two predictions on the type of disagreement. These are then used to analyse the likely areas of disagreement in the song 'Little bit O' Soul'; initially recorded by the Little Darlings in 1967, and more famously recorded later the same year by the Music Explosion, whose version went to no.2 on the *Billboard Hot 100*.

Overall, this chapter finds that Riemannian theory can at least partially explain 40% of the annotator harmonic disagreement. I find that disagreement on a recurring musical feature (such as a repeating motive in an instrument), and disagreement where a chord starts and finishes, are important causes of harmonic disagreement. Overall, though a score enables the exploration of musical disagreement not explainable by Riemannian theory (such as recurring musical features), the result of using Riemannian theory with (Chapters 3, 4, 6) or without the score (Chapter 5) appears comparable. Therefore, I conclude this thesis by recommending that a substitution-only version of Riemannian theory would be hugely beneficial to music similarity applications, while also enabling an easy computational implementation as it does not rely on computational score availability — which is limited.

6.1.1 Predictions

From the previous three chapters (Chapters 3, 4, 5) I make the following predictions about harmonic inter-annotator disagreement:

1. If the melodic line changes to a new harmony before or after the other (accompaniment) parts, disagreement will occur. For example, in Chapter 4, this caused some of the annotator disagreement in bar 59 of 'Hotel California' by the Eagles (Figure 4.3).

2. Disagreement that arises on a prominent musical feature will occur throughout the track (e.g. a prominent guitar riff or the setting of the words from the song's title); therefore, annotators will be consistent when annotating repetitions. For example, this was seen in the songs 'Super Freak' by Rick James in Chapter 3 and 'Take on Me' by A-ha in Chapter 4.

3. Disagreement is likely to occur where a chord starts or finishes; this was particularly prominent in the dataset used in Chapter 4.

4. Disagreement is most likely to occur at the beginning and/or end of phrases; this was a prominent observation in the dataset used in Chapter 5.

5. Disagreement arises from annotators transcribing the harmony at different levels of annotation granularity. For example, the title phrase 'Take on Me' in A-ha's song, in Chapter 4.

In addition to predicting what may cause harmonic disagreement to arise, this thesis' previous three chapters also present predictions on the type of harmonic disagreement:

1. The *Parallele* substitution will most commonly explain disagreement (as shown in Chapter 3, and the interview study of Chapter 4).

2. Participants who hold a music degree will have more of their disagreement explained by the `Partial Agreement` category than the `Agreement` category. The reverse will be the case for those without a music degree. This chapter follows the same definitions used for these categories in Chapters 3 and 4: `Agreement` refers to the chord-label disagreements on which there is a full agreement on the Riemannian function, and `Partial Agreement` refers to a majority agreement in the function between the unique substitutions. This prediction arises from a comparison between Chapters 3, and 4's results in Chapter 4's conclusion.

6.1.2 Dataset

Only one song was chosen for the dataset in this chapter, as I did not want the length of time required to complete the study to be a limiting factor in securing participants (previously in this thesis, the time commitment required from participants has been a deterrent). I predicted that annotating one song would take a participant on average 20–30 minutes.

Often, only pieces of well-known popular music have transcriptions available; therefore, I used the Billboard dataset discussed in Chapter 3,[2] as this contains the most popular songs in North America. Although I wanted to choose a piece popular enough to have an existing transcription, I also wanted to try to eliminate the effect of familiarity bias on the annotators' harmonic annotation. (If an annotator had played the piece before they may be able to remember the harmony, and therefore might not be transcribing what they hear, but instead transcribing what they know). Therefore, I removed songs performed by significant artists, such as Elton John, and The Beatles. Next, I searched the remaining songs on YouTube, and reduced the dataset to songs that had fewer than a million plays (at the time of writing). Following this, I checked whether the songs had transcriptions available on *Musicnotes*. Four songs passed these filtering stages: 'Little Old Lady from Pasadena' by Jan and Dean (1964), 'Dance Away' by Roxy Music (1979), 'Redneck Friend' by Jackson Browne, and 'Little bit O' Soul' by the Music Explosion (1967). I analysed these four songs, to see which song featured passages (or could be re-composed to have passages) of the sorts highlighted in Section 6.1.1. I also confirmed that the pieces conformed to popular music standards; for example, having a repeating structure with a chorus section and a harmonic structure based on

2. Burgoyne, Wild, and Fujinaga, "An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis."

FIGURE 6.1: Bar 12 of 'Little Bit O' Soul' as recorded for this study, showing the change in harmony from an F2 to an E2.

the Tonic–Subdominant–Dominant–Tonic progression detailed by Riemann. I chose the song 'Little bit O' Soul' by The Music Explosion. The song is in G major, and at the time of writing it had 748,428 plays on YouTube.[3]

To ensure that the song explored all the predictions made in Section 6.1.1 a few harmonic adjustments were made, including substituting a G major chord (the tonic) for an E minor chord (`Tp`) in bars 12 and 40, as shown in Figure 6.1. To, again, ensure that time commitment was not a deterrent, I reduced the song to 1 minute 35 seconds. Importantly for the predictions made in Section 6.1.1, I made sure to retain aspects of repetition and new material, and that the structure of my shortened version remained authentic to the structure of the full piece.

To ensure that the score aligned entirely with the audio recording, I had a recording made of the song using piano and voice (and added a sampled drum-kit in production). The performers were provided with the score, and instructed to play the piece as written,

3. ourFAMILYvideoLOG, 'A Little Bit Of Soul the Music Explosion', YouTube, 10 April 2009, https://www.youtube.com/watch?v=CgGjvZcNpKs [accessed 25th November 2019].

and, most importantly, I specified that they were not to change or deviate from the written harmony.

6.2 Methodology

Principally, this study is a harmonic transcription task, based on the work of previous harmonic annotation studies, such as Trevor de Clercq and David Temperly (2011), and Vincent Koops et al. (2019).[4] To focus the annotators' task, Koops et al. (2019) provided participants with a grid on which to annotate the harmony according to the suggested segments (annotators transcribed the harmony per beat).[5] By contrast, in this study, I wanted to ensure that the participants annotated the harmony according to their own perceptions of the harmonic changes, and providing them with a beat-segment grid could influence their thinking. Therefore, I provided the annotators with a rhythmic reduction of the vocal line and the associated lyrics (see Figure 6.2). This method is similar to that of de Clercq and Temperly (2011), where the authors annotated the pieces themselves on the score, and participants were consequently able to mark a harmonic change at any point on the score.[6]

Participants were asked a set of demographic questions before the transcription task, including their age (in a closed question format), musical performance qualifications (closed question), musical academic qualifications (closed question), instrument (specified response format), participation in ensembles (boolean answer format), listening habits (closed question), and experience in harmony annotation (open response format). These demographic questions were in the same format as the questions used in the second study in Chapter 4, and are also similar to the format used in Chapter 2.

For the transcription task, the participants were played 'Little Bit O' Soul', and then asked to transcribe the harmony on the provided rhythmic reduction. The participant was given free rein in terms of the length of time they took to complete the task, the methodology they used to transcribe the harmony, and how many times they wished to listen to the piece following its first play. Specifically, I asked the participants to:

> Annotate the chord sequence of the piece on the rhythmic reduction of the melody line. Please give chord labels for the harmony, using any method to determine the chord.

This study was limited in size because it was not possible to offer a financial incentive. This size limit meant I could not adopt a quantitative approach, which, would have

4. Clercq and Temperley, "A corpus analysis of rock harmony."; Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."

5. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."

6. Clercq and Temperley, "A corpus analysis of rock harmony."

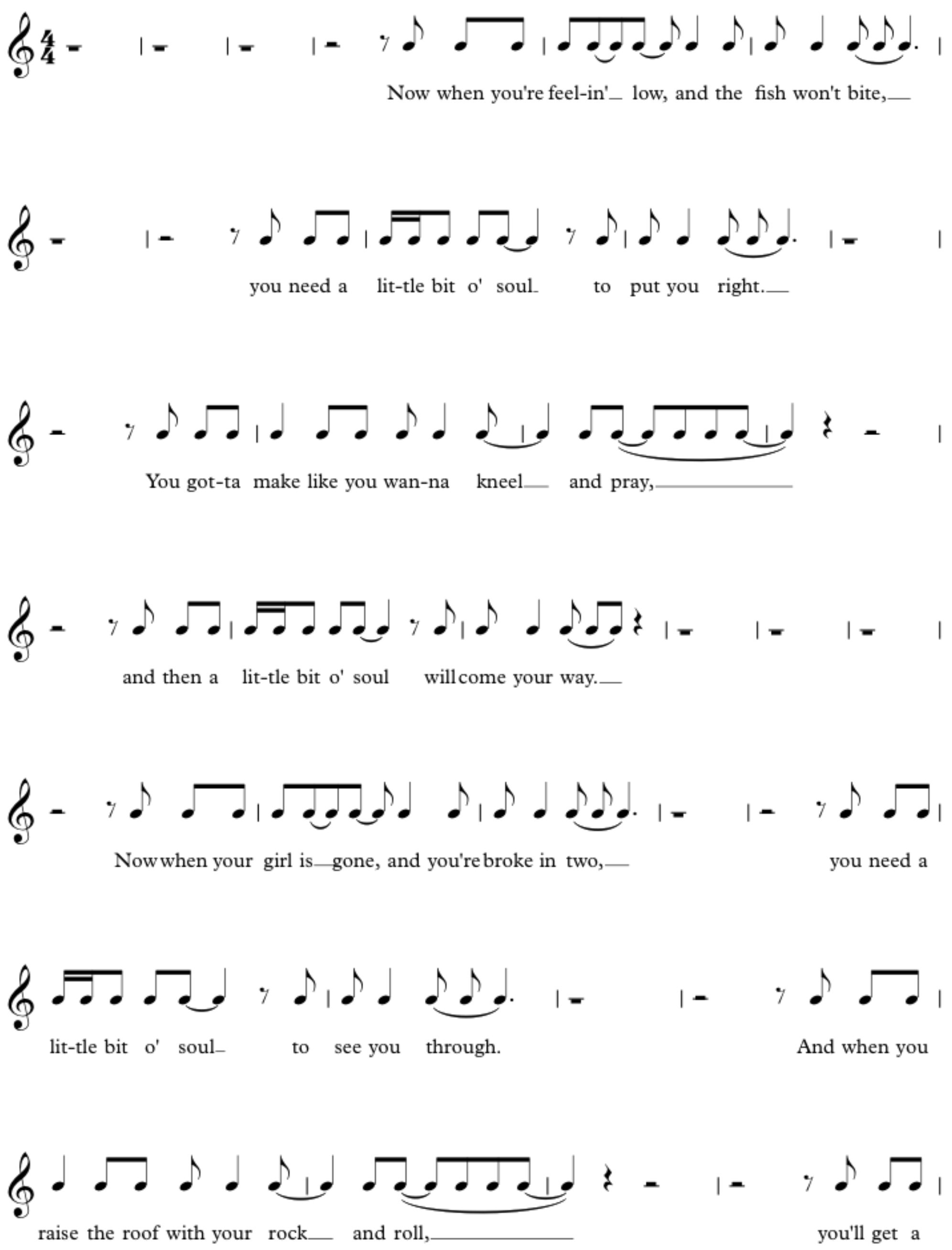


FIGURE 6.2: The rhythmic reduction given to annotators to label the chords of 'Little Bit O' Soul' (this is not the full song but the first page of a page-and-half reduction sheet).

enabled me to generalise the results to the general population. Instead, I decided to take a more qualitative approach, where I used participants' observations to test the validity of my predictions and inform avenues for further investigation. Therefore, I followed the transcription with a semi-structured interview, so that, as well as observing the methodology of the participant in transcribing the harmony, I was able to ask further follow-up questions. I asked three questions, allowing ample space for elaboration:[7]

1. Do you have any comments about the process?
2. Is there any part of the song that you found particularly difficult to transcribe?
3. Have you heard the song before? If you have, have you ever played it?

6.3 Participants

This study's participants consisted of 15 individuals with differing levels of musical experience.[8] The participants were chosen based on their academic music and performance experience. Participants had varying levels of academic music experience, including one participant (7%) with a PhD in music composition, 59% (9/15) of the participants with a masters in music (musicology, music theory or composition) and 7% (1/15) of the participants with a GCSE in music. The remaining 27% (4/15) of the participants had no formal academic music qualifications. The participants were mostly aged between 25–34 (9/15); see Table 6.1. The majority of participants listened to music for more than 15 hours per week (6/15); see Table 6.2. All 15 participants had experience playing in an ensemble. 27% of the participants (4/15) described themselves as popular musicians; meaning they either played popular music to a greater degree than classical music. In contrast, 53% (8/15) of the participants worked with or played classical music more prominently. 30% of the participants did not identify with either genre more prominently. Of the musicians (those with at least a masters in music), 40% (4/10) were composers, and another 40% (4/10) were music analysts; the remaining 20% identified themselves as neither. 40% of the participants had professional experience in harmonic annotation.

Of the participants, 27% (4/15) had no ABRSM graded or equivalent performance qualifications. The remaining 73% (11/15) of the participants had at least an ABRSM Grade 5 qualification. 53% (8/15) of the participants had an ABRSM Grade 8 or equivalent in performance. The participants in this study played a variety of instruments, as shown in Table 6.3.

7. Soafer, "Qualitative Methods: What are they and Why use them?"

8. Originally 16 participant's completed the study, but upon examination of one participants results, it appeared that they annotated the melodic line of the piece instead of the harmony of the song. As they completed the exercise incorrectly, I removed this participant's results from the study.

Age	**Frequency**
18–24	0%
25–34	60%
35–44	20%
45–54	13%
55–64	0%
65–74	7%
75-Plus	0%

TABLE 6.1: The proportion of participants that fell within the different age categories (these are the generic age categories often used in research studies). The total number of participants was 15.

Hours Listening Per Week	**Frequency**
None	0%
Less than 1	0%
1–3	13%
3–6	27%
6–10	7%
10–15	13%
15 Plus	40%

TABLE 6.2: The proportion of participants that fell within the different listening hour categories.

Instrument	**Frequency**
Flute	7%
Guitar	7%
Oboe	7%
Piano	53%
Violin	7%
Voice	13%
None	27%

TABLE 6.3: The proportion of participants who play each instrument listed in response to the study. The four participants said that they could not play an instrument ('none') were the same four who had no instrumental qualification, and therefore likely had no instrumental playing experience at all. In contrast, 27% participants (4/15) played multiple instruments.

6.4 Materials and Analysis

I present my analysis of my re-composition and recording of 'Little Bit O' Soul', performed according to the previously discussed five predictors as to where disagreement is likely to occur, and I will predict the sections of the song that will likely feature annotator disagreement. If these predictions do correctly identify areas and types of disagreement within this song, then it will be possible to provide reasons for some harmonic disagreement. In turn, these predictors may be able to provide further explanations for harmonic inter-annotator disagreement, which could prove useful in music similarity applications.

6.4.1 Prediction 1: Harmony changing at different times in different instruments

My first prediction states that if the melodic line moves to a pitch that is in a new harmony at a different time to the other (accompanying) parts, then disagreement will often occur. This prediction is of a similar nature to Predictions 3 and 4, which also discuss disagreement arising at points of harmonic change, or the beginnings and ends of phrases. Such disagreement is usually caused by different annotators attributing the harmonic change to different points, often in different instrumental parts.

In 'Little bit O' Soul', I predict this cause of disagreement will lead to annotator disagreement at the beginning of each verse: for example, bars 4–5 (see Figure 6.3). The harmony in these bars moves from 'N.C.' or 'No Chord', to G major (`T`). Arguably, the harmony does not begin until both the piano and vocal part play together (as prior to this the piece is monophonic) at the beginning of bar 5, before then moving to C major (`S`) halfway through the bar. However, I predict that some annotators will label the first chord at the beginning of the vocal line, to the words 'Now when you're', which serve as an upbeat in bar 4. Though there is no harmony in bar 4, melody could be seen as part of the harmony in bar 5, as these introductory words are sung on a D, which form part of the G major chord (G, B, D). I expect there to be a disagreement over this chord's placement between those participants who have an academic classical music background, and those who do not. The academic music convention of notating harmony on downbeats may lead those from such a background to keep their annotations to those downbeats, causing disagreement between annotators on the placement of chords; this will be explored further in the results section. Disagreements like this were particularly prevalent in Chapter 4, where bar 59 of the Eagles' 'Hotel California' saw the annotators disagree on whether the D major chord started on beat 1 of bar 59 with the bass line or on the second beat of the bar by which point all parts were playing a D major harmony.

6.4.2 Prediction 2: Disagreement on a prominent musical feature

My second prediction states that if a disagreement arises on a prominent musical feature (for example, a prominent guitar riff or the repetition of the words from the title of the song), this disagreement will repeat with each re-occurrence of the prominent musical feature. Recurring disagreement on prominent musical features explained a substantial amount of the disagreement in Rick James' 'Super Freak', discussed in Chapter 3, where a disagreement arose between the annotators regarding the harmony of the guitar riff.

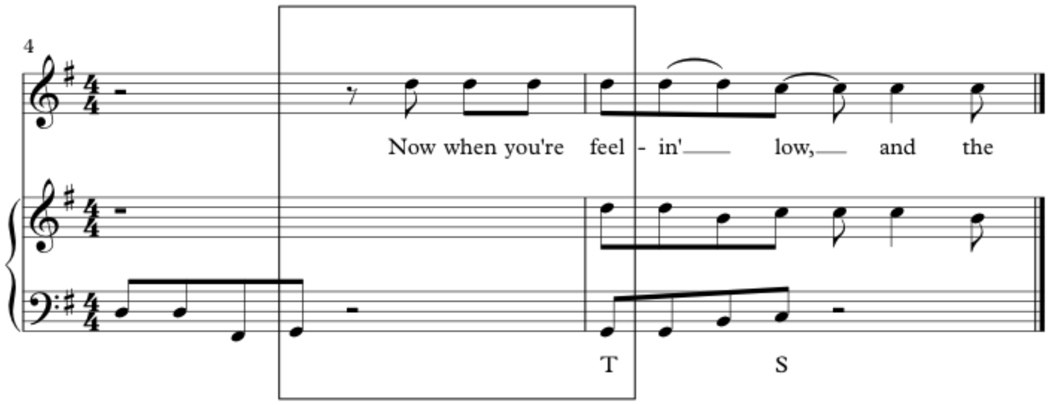


FIGURE 6.3: Bars 4–5 of 'Little Bit O' Soul', where participants disagreement over whether the harmony begins at the end of bar 4 or the beginning of bar 5 (highlighted).

This disagreement was then repeated continuously throughout the piece (as the guitar riff is featured continuously throughout the piece as part of the accompaniment). In total, this guitar riff disagreement caused 28% of the harmonic disagreement in this song.

'Little Bit O'Soul' has a repeated piano riff (in the original recording this is played by a guitar). This piano riff begins at the introduction (bars 1–4) and continues throughout the entire piece. What is particularly interesting is the similarity between this riff and the riff in 'Super Freak', as they both feature the 'N.C.' or 'No Chord' annotation when the riff is present without a fuller harmony (bars 1–4, bars 19–22 and bars 47–50 — see Figure 6.4). I, therefore, predict that if disagreement occurs over the harmony of this piano riff, then that disagreement is likely to continue throughout the song.

6.4.3 Prediction 3: Disagreement on where a chord starts or finishes

My third prediction states that disagreement will arise at points of harmonic change (i.e. chord changes). 15% of the harmonic disagreements in Chapter 4 were at points of harmonic change, but this could have been a product of the various beat-tracking algorithms used by Chordify. I will use this result from Chapter 4 as a prediction to see if other factors, such as the academic background of the participant, could explain disagreement to a greater or lesser degree. Such disagreement could arise in 'Little Bit O' Soul' in bars 8–9 (Figure 6.5) where there is a chord change from D major (D), to G major (T), to C major (S). Annotators could perceive the G major harmony as not starting until the beginning of bar 9 (on a strong beat), or they could perceive it as

FIGURE 6.4: The riff of 'Little Bit O' Soul', as it features in bars 1–4, 19–22 and 47–50, showing the development of the theme, and the filling out of the harmony.

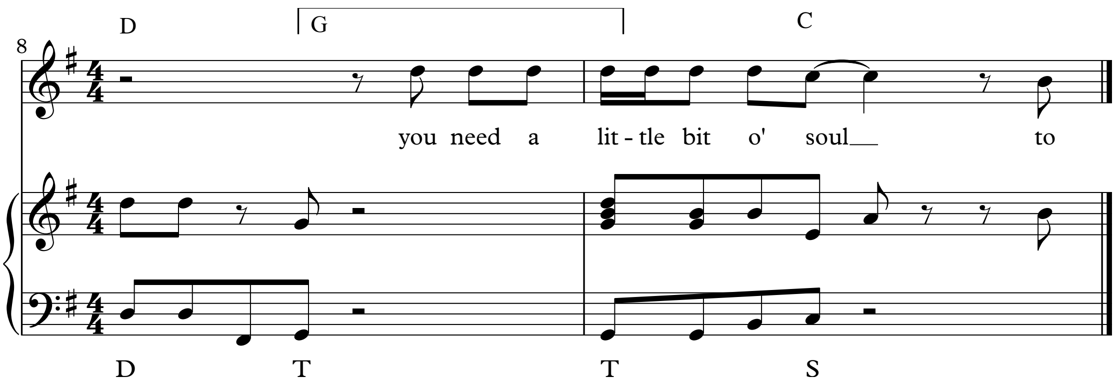


FIGURE 6.5: Bars 8-9 of 'Little Bit O' Soul', showing the possible disagreement over the positing of the chord G major with a bracket above the top stave.

starting with the lyrics (on a weak beat), or on the word 'need' or on the rest before 'you', as these are also strong beats. Though the annotators could disagree on the exact placement of the chord, either perceptually or due to notational convention, the annotators still agree (in this prediction) on the nature of the chord change.

6.4.4 Prediction 4: Disagreement at the beginning and/or end of phrases

My fourth prediction was borne out of the study in Chapter 5, where 43% of the harmonic disagreement occurred at the beginning or end of a phrase. By examining the lyrics of the song 'Little Bit O'Soul', we can predict where disagreement is likely to occur by looking at the first and last words of each phrase:

Now when you're feelin' low, and the fish won't bite,
you need a little bit o'soul to put you right.
You gotta make like you wanna kneel and pray,
and then a little bit o'soul will come your way.

Now when your girl is gone, and you're broke in two,
you need a little bit o'soul to see you through.
And when you raise the roof with your rock and roll,
you'll get a lot more kicks with a little bit o'soul.

And when your party fails 'cause ain't nobody groovin',
a little bit o'soul and it really starts movin',
yeah.

Just remember what I said 'bout a little bit o' soul.

We expect the disagreement to fall on the words 'now', 'bite', 'you', 'right', 'you', 'pray', 'and', and 'way' in the first verse. These fall in bars 4, 6, 8, 10, 12, 14–15, 16, and 18 of the full score. Similarly, in the rest of the song, we would anticipate the disagreements to occur on the first or last words of a phrase.

6.4.5 Prediction 5: Disagreement caused by differences in annotation granularity

My final prediction is that differences in annotation granularity will cause harmonic disagreement between participants. In Chapter 3, I explained 5% of the annotator disagreement by noting that the annotators transcribed the harmony at different levels of granularity. In Chapter 4, I observed that differences in annotation granularity explained 3% of the disagreement. This predictor relates to the concepts of Schenkerian prolongation,[9] and suggests that annotators may hear harmony changing at different rates (or in Schenkerian terms, we hear a piece at different levels of the hierarchical structure), and judge different changes as relevant or irrelevant to the change of the actual harmony.

Differences in annotation granularity are particularly likely in bars 12–13 of 'Little bit O' Soul'. Figure 6.6 shows the harmony of bar 12 changing from D major, to E minor, to G major in one bar, with the G major continuing into bar 13. However, the E minor may be regarded as a passing chord, as it occurs nowhere else in the harmony (previously, in places such as bar 8, the harmony just moved from D major to G major). Therefore, annotators could perceive this harmonic change as an elaboration of the existing fundamental harmonic change of D major to G major. Additionally, E minor is the *Parallele* substitution of G major and therefore harmonically close to the chord it is substituting in the elaboration. We can, therefore, predict that an annotator who annotates at a more granular level would note the E minor chord, but one who annotates at a less granular level would label only the D major to G major harmonic progression.

9. Cadwallader and Gagne, *Analysis of Tonal Music: A Schenkerian Approach*; Larson, "The Problem of Prolongation in 'Tonal' Music: Terminology, Perception, and Expressive Meaning."; Forte and Gilbert, *Introduction to Schenkerian Analysis*.

FIGURE 6.6: Bars 12–13 of 'Little Bit O' Soul' showing the possible disagreement over whether to annotate Tp (E minor) in Bar 12.

6.5 Results

In total 35% of the disagreement between annotators is at least partially explainable in this study using Riemann's theory of harmonic functions. Of this, 27% of the disagreement was of the category `Partial Agreement`, and 8% in the category `Agreement`. Annotator A_2 was responsible for the most disagreements: 14%, which is double the percentage of disagreements caused by any other annotators — A_{10} (the next highest annotator to disagree) caused just 6% of the disagreements. A_2's disagreement was 3.4 standard deviations from the mean percentage of disagreement caused by a single annotator, and 6.6 standard deviations away from the mean disagreement caused by all annotators excluding A_2. Interestingly, A_2 was a classical musician from a music-theoretical background. 38% of the time (20/52 bars) they annotated only a single chord per bar, and at most they annotated two chords per bar (62%, 32/52). As this piece uses a duple metre (such as 4/4), A_2 only seems to annotate on the two strongest beats, beat 1 and beat 3. I have, therefore, chosen to examine the agreement among the remaining 14 annotators, excluding A_2, to remove any variability and statistical power this outlier could have. As shown in Table 6.4, the percentage of disagreement that is at least partially explainable increases to 40% when removing A_2.

Overall, the majority of the annotator disagreement is explainable by the *Variante* substitution when excluding A_2. This is dissimilar to previous chapters, where in Chapters 3 and 4 I found that the *Parallele* substitution explains the highest amount of disagreement, and in Chapter 5 it is the *Leitonswechsel*. In this study, as with Chapter 5, the *Parallele* substitution was the second most common substitution to explain disagreement (10% of the disagreement). Interestingly, in Chapter 3 (where Prediction 6.4.2 originated from), all the participants held an academic music qualification. In comparison to this chapter (in Table 6.4), the percentage of disagreement explained by the

Parallele substitution is higher in categories where participants held a music degree, such as classical musicians and analysts who come from a traditional music education background, suggesting that the *Parallele* substitution is more likely to explain annotator disagreement among those with an academic music qualification. Interestingly, the *Leittonswechsel* explained 100% of the disagreement among non-musicians.[10]

Table 6.4 shows the different categories of `Agreement`, `Partial Agreement` and `No Agreement`, and the different substitutions used in explaining disagreement — *Parallele*, *Variante*, and *Leittonswechsel* — per demographic group. In the demographic categories for either having or not having a music degree, we see no difference in the percentage that can be at least partially explained by Riemannian theory (22% for the no music degree category, and 22% for the music degree category). The division between `Agreement` and `Partial Agreement` categories is also not substantial. For those with a music degree, 10% fall in the `Agreement` category, and 12% in the `Partial Agreement` category. For those without a music degree, 11% fall in the `Agreement` category, and 11% in the `Partial Agreement` category. The main difference between these two demographic groups is the different substitutions that explain disagreement. 100% of the disagreement between participants with a music degree can be explained using the *Variante* substitution, whereas, 100% of the disagreement between participants without a music degree can be explained using the *Leittonswechsel* substitution. The *Leitonswechsel* substitution was also important in explaining the disagreements between participants in Chapter 3 and contrastingly, the participants in Chapter 3 all had a music degree, and therefore belonged to the music degree demographic category.

In the 'musicians' category, I sub-divided the demographic group into 'popular' and 'classical' musicians, to see if disagreement was affected by participants' musical style specialisms.[11] Riemannian's theory is a music theory written for classical music, though prominent discourse argues in favour of its use in popular music.[12] Therefore, I also designed these sub-categories to explore whether 'classical' musicians' disagreements would be more explainable by this approach. The 'popular' musicians' disagreement was only explainable in the `Agreement` category (22%), they had no annotator disagreement in the `Partial Agreement` category. In contrast, the 'classical' musicians had some disagreement explainable by the `Agreement` category (8%), but the majority of their disagreement was explainable by the `Partial Agreement` category (21%). Overall, only 13% more of 'classical' musicians' disagreements were at least partially explainable using Riemann theory.

This comparison between musicians and non-musicians builds on the discussion in Section 4.4 of Chapter 4, which suggested that a higher level of disagreement is explainable

10. For this results section I define non-musicians as participants without an academic music qualification.
11. Inverted commas used to distinguish the demographic group from the general discussion.
12. Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music"; Biamonte, "Modal Function in Rock and Heavy Metal Music."

Category	**All**	**No A2**	**No Degree**	**Music degree**	**Pop**	**Classical**	**Composer**	**Analyst**
`Agreement`	8%	17%	11%	10%	22%	8%	21%	8%
`Partial Agreement`	27%	23%	11%	12%	0%	21%	0%	2%
`No Agreement`	65%	60%	78%	78%	76%	71%	79%	90%
Parallele	0%	10%	0%	0%	0%	22%	0%	55%
Variante	100%	90%	0%	100%	100%	78%	100%	0%
Leittonwechsel	0%	0%	100%	0%	0%	0%	0%	45%

TABLE 6.4: The proportion of disagreement according to the different agreement categories `Agreement`, `Partial Agreement` and `No Agreement`, as divided by different demographic groups. The second half of the table shows the percentage of each Riemannian substitution that explains the disagreement in the category `Agreement` for the different demographics. The demographic groups are divided into: all participants (15), all participants removing A2 (14), participants without a music degree (5), participants with a music degree (10), participants who stated popular music as their genre of work (4), participants who stated classical music as their genre of work (8), composers (4) and music analysts (4).

by the category `Partial Agreement` for musicians. This chapter's study (Chapter 6) saw that those with a music degree had slightly more (12% vs 10%) of their disagreement explained by the `Partial Agreement` category rather than the `Agreement` category. However, the demographic category for those without a music degree showed no difference in the percentage explainable by the categories `Agreement` and `No Agreement`. Therefore, this chapter's results suggest the difference in terms of which agreement category explained annotator disagreement has more to do with the musical specialism of the annotators rather than whether or not they have a music degree. Chapter 3's participants featured classically trained musicians prominently, whereas Chapter 4 prominently featured musicians interested in popular music. In this chapter, a more substantial proportion of disagreement was explainable in the `Partial Agreement` category for classical musicians (21% vs 8%). In contrast, all of the popular musicians' disagreement was explainable by the `Agreement` category. This aligns with the findings in Chapters 3 and 4, which showed a higher percentage of disagreement explained by the `Partial Agreement` category in Chapter 3 than Chapter 4. In turn, this suggests that it is not just those with higher music vocabularies that are more likely to disagree on harmony, but that those who specialise in music analysis are more likely to disagree on harmony — suggesting either a higher auditory agreement between popular musicians, or that a lack of domain knowledge (this study used a popular piece) led to greater disagreements.

The remainder of this results section will discuss the five predictions I made and whether or not they successfully predicted disagreement. I will also discuss the post-transcription interviews. Overall, the results indicate that predictions 2 and 3 were the most accurate, whereas Prediction 5 (annotation granularity) only explained A_2's disagreement.

6.5.1 Prediction 1: Harmony changing at different times in different instruments

This chapter's first prediction stated that harmony changing at different points in different instruments would lead to disagreement. In total, this explained 10% of the annotators' disagreement, and 6% after removing A_2. Prediction 1 equally explained the disagreement between those with and without music degrees (both 7%). Interestingly, none of the 'popular' musicians' or composers' disagreements were explainable using this method. In contrast, 10% of the disagreement between annotators in the 'classical' sub-genre was explainable using this method, and 15% of those in the music theorist demographic group.

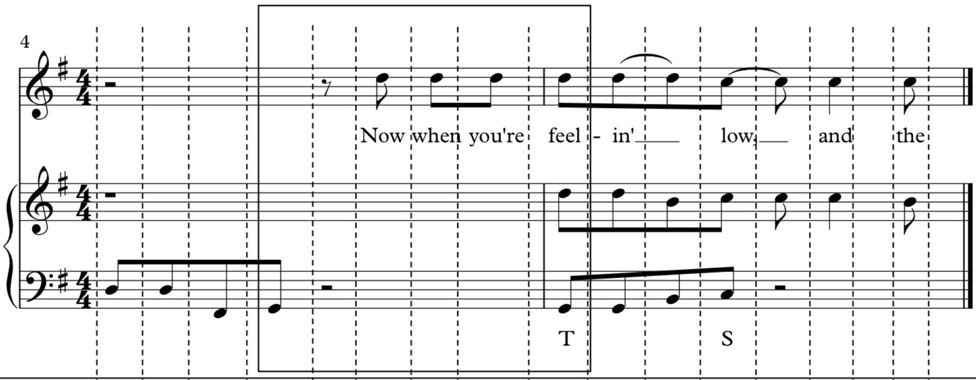
Figure 6.7 shows the different annotators agreeing on the overall harmony but disagreeing on where it starts in bars 4–5. A_1 perceived no harmony until the piano and vocal parts fill out in bar 5. In contrast, the most common transcription agreement is the change to G major (T) at 'sub-beat' 4 of bar one, at the pitch G in the piano part. However, not all of the annotators agree on this: A_5, A_{13}, and A_{14} see the change on 'sub-beat 5', and A_{15} on 'sub-beat 6'. The vocal part begins on beats 5 and 6, so the annotators may have chosen 'sub-beat 5' as the strong beat (beat 3 and 3/4 time), rather than 'sub-beat' 6 which introduces the new harmony.

6.5.2 Prediction 2: Disagreement on a prominent musical feature

This second prediction expected disagreement that occurred on a prominent musical feature would recur when this feature returned throughout the piece. In this study, 19% of the disagreements between all the participants are explainable using this method. What is particularly interesting is that if we remove A_2 (the annotator shown earlier to be the most likely to disagree) from the results, this drops to 6%. Therefore, 13% of A_2's disagreement not explainable by Riemannian theory was due to a recurring disagreement on a prominent musical feature. In this case, the feature was the piano riff discussed in Section 6.4.2. As A_2 caused over 13% of the whole population of explainable disagreement for this prediction, it is important to remove A_2 and observe the amounts explainable without this anomalous contributor.

Figure 6.8 shows the repeating piano riff that forms the basis of the piece. Disagreement arises in this passage at points where the harmony changes. The chord sequence is `T`–`S`–`D`–`T` or `I`–`IV`–`V`–`I`. The different annotators perceive this change at different points of the bar. Annotators A_1 and A_{15} disagree, as they see this passage as having 'No Chords' ('N.C.'), as it is a single piano line. The only annotators that disagree on this harmony are A_2 and A_{10}, who perceive the harmony as `T`–`D`–`T` or `I`–`V`–`I` instead.

Overall, looking at the demographic categories, 'no music degree', 'popular musicians', and 'composers' all had 0% of their disagreement explained as disagreements arising on a crucial musical feature; interestingly, none of these categories include A_2. This prediction explains 16% of the disagreement between those with music degrees, 18% of the disagreement between classical musicians and 28% of the disagreement between music theorists.



A_1	N.C	*	*	*	*	*	*	*	G:maj (T)	*	*	C:maj (S)	*	*	*	*	*
A_2	D:maj (D)	*	*	*	*	*	*	*	G:maj (T)	*	*	C:maj (S)	*	*	*	*	*
A_3	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_4	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_5	D:maj (D)	*	*	*	G:maj (T)	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_6	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_7	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_8	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_9	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_{10}	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	*	*	*	*	*	*
A_{11}	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*	*
A_{12}	D:maj (D)	*	*	*	*	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_{13}	D:maj (D)	*	*	*	G:maj (T)	*	*	*	*	*	*	*	*	C:maj (S)	*	*	*
A_{14}	D:maj (D)	*	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_{15}	N.C	*	*	*	*	G:maj (T)	G:maj (T)	*	*	*	*	C:maj (S)	*	*	*	*	*
Disagreement Category			N	N	N	N	N	N	N								

FIGURE 6.7: Bars 4–5 of 'Little Bit O' Soul', showing annotator disagreement arising, as predicted, at the G major chord change at the end of Bar 4.

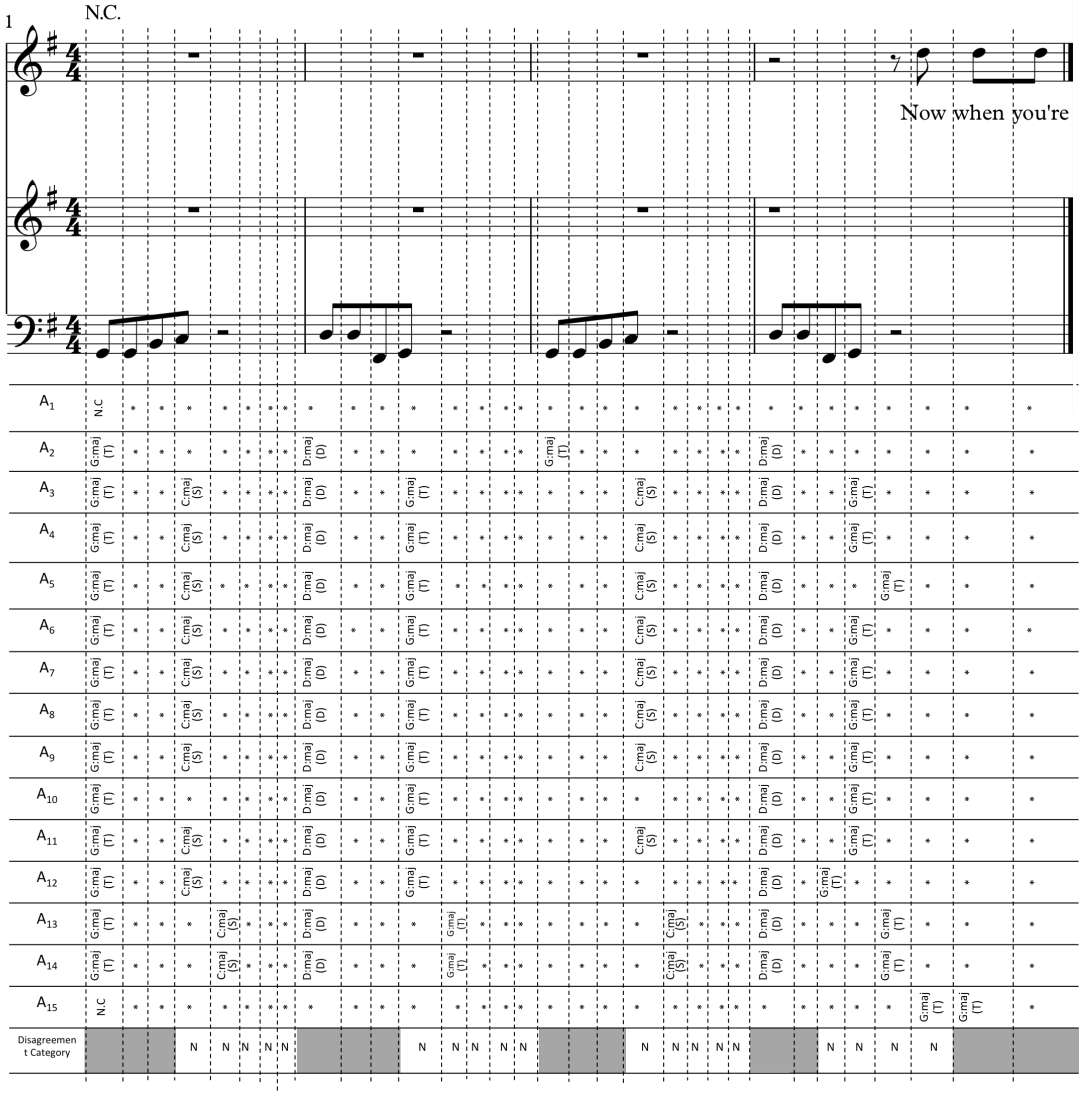


FIGURE 6.8: Bars 1–4 of 'Little Bit O' Soul', showing the disagreement that is repeated with the return of the piano riff.

6.5.3 Prediction 3: Disagreement on where a chord starts or finishes

Prediction 3 stated that disagreement was likely to occur at points where a chord starts or finishes. In this study, 29% of the annotator disagreement that cannot be explained by Riemannian theory occurs at points of harmonic change. Upon removing A_2, 39% of the disagreement that Riemannian theory could not explain occurred at these points. As predicted in Section 6.4.3, we can see this disagreement in bars 8–9. The harmonic sequence of this piece overall is `I–IV–V–I`. Figure 6.9 (bars 8–9) shows the end of the sequence and its recommencement `V–I–IV`. As can be seen, the annotators agree on the chord progression D major – G major – C major, but they disagree on whether the harmony changes on strong or weak beats throughout the bar. Most of the annotators perceive this change as being on the weak beat at the end of the piano riff; however, some annotators (3 and 4) wait for the next strong beat (beat 3), and some wait as far as the beginning of the next bar (1 and 2).

A_1	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_2	D:maj (D)	*	*	*	*	*	*	*	G:maj (T)	*	*	*	*	*	*	*
A_3	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_4	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_5	D:maj (D)	*	*	*	G:maj (T)	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_6	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_7	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_8	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_9	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_{10}	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	*	*	*	*	*
A_{11}	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_{12}	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
A_{13}	D:maj (D)	*	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*
A_{14}	D:maj (D)	*	*	*	*	G:maj (T)	*	*	*	*	*	*	C:maj (S)	*	*	*
A_{15}	D:maj (D)	*	*	G:maj (T)	*	*	*	*	*	*	*	C:maj (S)	*	*	*	*
Disagreement Category				N	N	N	N	N	N	N		N	N	N	N	N

FIGURE 6.9: Bars 8-9 of 'Little Bit O' Soul', showing disagreement on where the harmony changes from D major to G major, and again from G major to C major.

Of the disagreement categories, those with no music degree had the most considerable portion of disagreement explained by this prediction (59%), followed by music theorists (32%), classical musicians (24%), popular musicians (19%), composers (15%), and finally those with a music degree (9%). Interestingly, those with a classical music background and those with a strong theoretical background disagreed more often on the placement of a harmonic change, suggesting that the convention of notating the harmony only on strong beats was not adhered to in this task by these groups.

Removing A_2 from these demographic categories only sees a substantial change in how much of the music theorists' disagreements can be explained by occurring over the beginning or end of a chord, and raises the amount of disagreement explainable using this approach from 59% to 64%.

6.5.4 Prediction 4: Disagreement at the beginning and/or end of phrases

Prediction 4 suggests that disagreement will arise at the beginning or end of a phrase, as discussed in Section 6.4.4. Overall, 13% of the disagreements in this song fall into this category. In Chapter 5, 43% of disagreements were explainable using this same method. In this chapter (Chapter 6), this type of disagreement was particularly prevalent: for example in bars 6 on the word 'bite' (end of line 1), 8 on the word 'you' (beginning of line two), and 10 (end of line two).

This method was particularly important in explaining the disagreement between music analysts in this study (when removing A_2), explaining 28% of the disagreement (only 7% when A_2 was included). Interestingly, A_2 made little difference to the overall proportion of the disagreement explainable for the whole population: 13% of disagreement was explainable including A_2 and 12% excluding them. This method for explaining disagreement was not particularly important for any of the other demographic groups.

6.5.5 Prediction 5: Disagreement caused by annotation granularity

Prediction 5 was developed in response to the study in Chapter 3, which showed that we could account for 5% of disagreement through differences in annotation granularity. As noted before, A_2 disagreed most frequently with the other annotators in this study. It is particularly interesting that A_2's disagreement was consistently caused by them using a different annotation granularity to the other participants. Most annotators annotated four chords per bar (as the piece uses a duple metre), whereas, A_2 annotated only one chord per bar 38% of the time (20/52 bars), and two chords per bar for the remaining bars (62%, 32/52). In this study, none of the disagreement across any demographic group is explainable as a disagreement arising from differing levels of granularity at which annotators chose to annotate.

6.5.6 Follow-up Interviews

Following completion of the annotation task, participants were asked the following questions:

1. Do you have any comments about the process?
2. Is there any part of the song that you found particularly difficult to transcribe?
3. Have you heard the song before? If you have, have you ever played it?

The discussion in these interviews mostly focused on the methodology that each participant used to transcribe the harmony. Nine (9/15) of the participants stated they used an instrument in transcribing the harmony; either the piano or guitar. They used this instrument to either 'check' or 'confirm' the annotation they had completed by ear, or to 'recreate' the music. Five of the participants noted that they first identified the key and then transcribed the harmony in terms of the chords' Roman numerals and their relation to this identified key; of these participants, three (3/5) were popular music composers. Two of the participants, both popular music composers, used a computer or computational software such as Logic to record and loop back parts of their annotation.

A discussion with A_{15} about the influence that traditional popular music form had on their annotation was particularly interesting. They highlighted that they had expected that there would be a distinct 'middle 8' section in the piece, due to its existence in most popular music songs; specifically, they identified that a lack of this section confused them. In contrast, A_6 identified the harmonic change of bars 38–46 as a middle 8, as it features a move to the dominant D major, which is particularly noticeable due to the presence of the chord A major. Similarly to A_{15}, A_6 noted that they were looking for this section due to its prevalence in popular music, this emphasised the importance of form in annotators' transcription methodologies. It appears that participants have preconceived notions of the piece's form on the basis of its genre, and this shaped how they listened to and identified the harmony of sections.

6.6 Discussion and Conclusions

Overall, 40% of the disagreements in this chapter's study are explainable at least partially using Riemannian theory. This is comparable to the disagreement explainable in Chapters 3 and 4: 40% explainable in Chapter 3, and 50% in Chapter 4. Chapter 5 showed a comparable level of explainable disagreement, at around 41%, between seven songs. This chapter, therefore, suggests that the methodology proposed in Chapter 5 to remove the requirement of a score in analysis, and the one used in this chapter that utilises a score of the song are comparable in explaining disagreement.

In Chapter 3, 40% of the disagreement was explainable using Riemannian theory, compared to 50% explained in Chapter 4. Interestingly, the most significant difference was in the percentages explainable by the categories of `Agreement` and `Partial Agreement`. The conclusion section of Chapter 4 noted that higher levels of disagreement were explainable in the category of `Agreement` when annotators were not 'expert' transcribers or annotators.

This chapter shows that among participants with a music degree, slightly more disagreement is explainable by the category `Partial Agreement`. Interestingly, however, all four participants in the interview in Chapter 4 came from a popular music background. Looking at the results of this chapter's study, all of the disagreement between participants from a popular music background falls into the category `Agreement`; by comparison, the classical musicians had a higher amount of disagreement explained in the category of `Partial Agreement`. Comparably, the same is the case between the composer and music analyst categories. It is worth noting here that the composers in this study were all from a popular music background, meaning that we cannot differentiate between a composer and a popular musician, or indeed between an analyst and a classical musician, as in this study they are the same annotators. This suggests that this difference is not merely caused by whether or not the annotator has a music degree. Instead, the results of this chapter allude to the principal demographic difference being whether the participant's interest in chord transcription comes from a popular or classical background, or a composition or analytical background. Therefore, we cannot conclude whether it is the style of music that the participant is interested in, or the type of musical activity (composer or analyst) that accounts for this difference between the agreement categories.

I predicted that the *Parallele* substitution would be the most common explainer of disagreement in this chapter; this prediction did not hold true for this dataset and these annotators. Instead *Variante* substitution explained the most substantial proportion of disagreement, most significantly for those with a music degree, popular musicians and composers. However, if we remove A_2 from the dataset, we can see that the *Parallele* substitution does explain some of the annotator disagreement: 22% of classical musicians' disagreement, and 54% of analysts' disagreement. Though the *Parallele* substitution did not explain the most disagreement in this dataset, it is important to note that, since this dataset was limited to one song, and consisted of a relatively small number of participants, we do not know if this is an anomalous result due to this studies small nature, or if there is a more complex relationship between demographic features and the type of substitutions that explain the most disagreement.

This chapter introduced a variety of predictions in terms of both where and how participants were likely to disagree; see Section 6.1.1. The first set of predictions, concerning where disagreement would arise, saw prediction 2 (disagreement on a prominent musical feature) and 3 (disagreement on where a chord starts and finishes) explain the most disagreement: Prediction 2 explained 19% of all the participants' disagreement, and prediction 3 explained 39% of all the participants disagreement. Prediction 2, in particular, explained all of A_2's disagreement. Overall, prediction 3 (disagreement on where a chord starts and finishes) explained the most considerable amount of participant disagreement (39%). This prediction was particularly prominent in explaining disagreement between those with no academic music qualifications (59%) and music theorists (32%).

This chapter has raised possible future avenues for further investigation, including the influence of musical background (popular vs classical, and music analyst vs composer) on both the amount of disagreement explainable using Riemannian theory, and the proportion that falls into the category `Agreement` versus the category `Partial Agreement`.

The methodology discussed in this chapter requires a musical score (i.e. a transcription) to carry out a Riemannian analysis. As discussed in both Chapters 3 and 4, using a musical score can be subjective. I attempted to negate this limitation from this chapter by using the score to record the piece, and creating an accurate representation of the score in audio format. However, many popular songs do not exist in a score format before being recorded; as discussed in Chapter 5, many popular musicians do not read sheet music, and they often compose without the use of a score.

Through the use of a score that existed prior to recording, this chapter has shown that Hugo Riemann's theory of harmonic functions can be employed without a musical score to explain annotator disagreement. The harmonic function aspect of the theory, enabled through having an accurate transcription of the score (identification of local and global key changes), is not imperative to the explanation of annotator disagreement, and thus music perception. Indeed, the use of substitutions alone as proposed in Chapter 5 explained a larger proportion (43%) of annotator disagreement than the approach taken in this study (in this study 40% was explained). Throughout this thesis, Riemannian theory has explained, on average, between 40 and 50% of the disagreement across the datasets. Alongside this, Chapters 3, 4 and 5, and this chapter, have shown that disagreement on prominent musical features, and at points of harmonic change requires further investigation in the future to see how a non-score-based approach could use these findings to improve harmonic annotation software, and music information retrieval techniques. This thesis has shown that there is a consistent level of disagreement that is explainable using Riemannian theory across a variety of demographic groups and popular musical examples.

Chapter 7

Discussions and Conclusions

7.1 Key Findings

Applications of music similarity are yet to utilise the potential of music theory in defining musical similarity. Perceptions of music similarity can be subjective as people hear similarity differently, and therefore it can be problematic to try to devise an effective means of detecting music similarity. The added musical knowledge gained from formal music theories could aid determinations of music similarity, music perception, music information retrieval, commercial music sales, and copyright law. As explored in this thesis, one such theory — Hugo Riemann's theory of harmonic functions — can explain inter-annotator harmonic disagreement, which in turn enables us to bypass issues of subjectivity when attempting to determine music similarity. This suggests that Riemannian theory can be used to explore and explain audible music similarity.

This thesis hypothesised that if one annotator can perceive a different chord to another annotator, then the two chords could be seen as perceptually similar, as they are audibly mistakable. This thesis used this hypothesis to explore harmonic similarity through a set of harmonic transcription and annotation studies (Chapters 3, 4 and 6). These studies explored which chords different annotators disagree on, and how Hugo Riemann's theory of harmonic functions might explain this disagreement. Chapter 3 found that 48% of the harmonic disagreements in the Chordify Annotator Subjectivity Dataset (CASD) were explainable using a music-theoretical approach. Of this 48%, Riemannian theory explained in full 27% of the disagreements between annotators, and a further 13% partially. I supplemented this approach by utilising other information from the musical score, enabling the explanation of a further 5% through disagreements caused by differences in the awareness of prolongation (a concept from Schenkerian analysis), and a further 3% through harmonic ambiguity. Overall, Chapter 3 determined that music theory can explain some harmonic inter-annotator disagreements, showing a higher level of agreement between annotators at this more musically informed harmonic function level. Perhaps most interestingly in Chapter 3, were comparisons between current MIR pitch-class agreements and this more musically informed harmonic function agreement. I noted that the chord labels `C:sus4` and `A:min` have no root note agreement, and no agreement on the root and third using the common MIREX evaluation measures. However, if analysed in the key of `C`, Riemannian theory reveals that these differing chords both fulfil a tonic function (as `T` and `Tp`, respectively). Vincent Koops et al. (2019) detailed a 69% agreement between the annotators using the common MIREX evaluation measures, whereas using a harmonic function analysis we can explain a further 16% of disagreement in this dataset (totalling 82% in this thesis's methodology compared to the 69% detailed by H. Vincent Koops et al. (2019)).[1] In turn, this directly shows how the creation of metrics that take into account the function of a chord in a tonal centre could provide a more nuanced view of chordal agreement and similarity.

The main limitation of the study discussed in Chapter 3 was the small sample size: just four annotators. Therefore, Chapter 4 used a larger dataset (a subset of the user edit data on Chordify). This crowdsourced dataset featured a more substantial number of participants; in total, 77 participants across 41 different songs. To date, this dataset is by far the most significant number of annotators that a study of harmonic inter-annotator disagreement has observed.

As I had only employed a quantitative methodology up to this point in this thesis, the second half of Chapter 4 detailed a qualitative interview study using the participants from the crowdsourced dataset. This allowed me to enrich the data collected, to verify or reject the quantitative results collected so far. I explored how and why participants made certain annotation decisions, raising, in particular, the importance of annotator granularity and points of harmonic change as potential causes of harmonic disagreement.

The methodology used in Chapters 3 and 4 required a musical score to perform a Riemannian analysis, which enabled the analyst to identify any local and global key changes. Due to compositional and recording practices in popular music, the available scores were often (subjective) transcriptions, if they were available at all. The lack of available scores for the songs in Chapters 3 and 4 was therefore problematic and led to a reduction of the datasets. Therefore, Chapter 5 proposed an approach to Riemannian theory that removed the functional element, but retained the notion of chordal substitution, which meant that I did not require a score for the analysis (as I did not need to identify the local and/or global keys). Instead, my analysis observed the relationship between the chords involved in disagreement. Overall, Riemannian substitutions explained at least some of the 40% harmonic disagreement in this chapter; this is similar to the 40% I explained at least partially using Riemannian substitutions in Chapter 3 (relating to the CASD dataset), and the 50% explained in Chapter 4. A further 5% of the harmonic disagreement in Chapter 5 occurred at points of harmonic change, suggesting that disagreements where precisely the harmony changes. In conclusion, Chapter 5 indicates

1. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."

that the harmonic function (and thus the key) is perhaps not important for determining similarity; only the concept of substitution is necessary.

Chapter 5 suggested that both a traditional Riemannian approach and an adapted approach (that does not require the score) equally explain annotator disagreement. However, up to this point, the scores used in this thesis had been (subjective) transcriptions. Therefore, Chapter 6 used a single song recorded from a score to ensure the accuracy of the score in relation to the audio file. I was, therefore, able to observe whether a traditional Riemannian approach, using the musical score, had any impact on the methods, ability to explain disagreement. The results of Chapters 3, 4, and 5 led to the formulation of a set of five predictions as to the likely areas and causes of disagreement in the song 'Little Bit O' Soul', originally recorded by the Little Darlings. This study found that Riemannian theory could at least partially explain 40% of the annotator disagreement (17% fully explained, and 23% partially). This is comparable to the disagreement explainable in Chapters 3 (40%), 4 (50%), and 5 (40%) (see Table 7.1 for an overview of the annotation studies, results in this thesis). Thus, Chapter 6 concluded that a score-based (traditional) approach and my adapted approach to Riemannian theory were comparable in their ability to explain harmonic disagreement among annotators.

Chapt.	**Dataset**	**No. Annotators**	**No. Songs**	**Explainable disagreement**	**Agreement**	**Partial Agreement**	**Prolongation**
3	Chordify Annotator Subjectivity Dataset	4	41	48%	27%	13%	5%
4	Chordify user edit data	77 (at most 11 per song)	41	68%	48%	2%	3%
5	Songs with no available scores from Chapters 3 and 4	At most 4 per song	7	45%	26%	14%	—
6	'Little Bit O' Soul'	15	1	%	17%	23%	0%

TABLE 7.1: Comparison of the results of my thesis's four annotation studies.

7.2 Other findings

7.2.1 Riemann's theory of harmonic functions explains a higher proportion of the disagreement among composers and popular musicians

Demographic features were explored in an attempt to explain the differing proportion of disagreement that was categorised in the `Agreement` and `Partial Agreement` categories in Chapters 3 and 4. Chapter 3 found that 27% of the explainable disagreement fell into the category `Agreement` and 13% in to the category `Partial Agreement`. In Chapter 4, a much larger percentage of disagreement was explainable by the category `Agreement` (48%) and only 2% was explained by the `Partial Agreement` category. Chapter 4 highlighted that this difference might be related to the music education of the participant, as the participants in Chapter 3 were expert transcribers, and those in Chapter 4 were not. Therefore, these results suggest that those with more extensive classical/theoretical harmonic training are more likely to have higher levels of disagreement, including on a functional level.

Chapter 6 further explored this, and concluded that the genre of the participant's musical background was potentially important, as popular musicians more commonly agreed on the function of the chord than classical musicians did. Similarly, composers were more likely to agree on the function of the chord than analysts were. It is worth noting that an agreement does not necessarily mean that the harmony is correct (this thesis did not study whether or not annotations were correct but instead observed disagreement). Importantly, this study cannot conclude whether it was both or one of these demographic features that explained this difference in the amount of disagreement explainable by the categories `Agreement` and `Partial Agreement`, as the participants in this study were mostly popular musicians who are also composers, and classical musicians who are also analysts. Similarly, it would be interesting to compare these demographic categories using classical works; it could be that the results are a by-product of the annotator being less familiar with the genre they were annotating. It would be interesting to know if the opposite would be true if the participants annotated the harmony for classical works: i.e. that classical analysts would be more likely to agree.

This increased explainability of popular musicians'/composers' disagreement is interesting, as we often see Riemannian theory taught in classical training and music theory, and not in popular music classes. However, jazz and popular music analysts have previously seen the relevance of a functional theory in analysing their genre.[2] The perceived audibility of the functional role of chords in popular music in this thesis could add to

2. Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music"; Everett, "Making Sense of Rock's Tonal Systems"; Doll, *Listening to Rock Harmony*.

the argument for the applicability of this theory in popular music analysis, and the importance of the theory in composition and music perception. Further research should investigate the usefulness of Riemannian theory in explaining harmonic disagreement in classical music, since this genre was used to develop this theory (Chapter 2 suggests that there may be an importance of Riemannian theory in classical music similarity, however, as that chapter's study did not explore annotator disagreement, this needs to be further investigated).

7.2.2 The *Parallele* substitution was the most common relationship between chords in harmonic disagreement

To complete the analysis of annotator disagreement in Chapters 3 to 6, I labelled sections of inter-annotator disagreement with Riemannian functions. Following this, the disagreement was categorised as being an `Agreement` in terms of the harmonic function a `Partial Agreement` or `No Agreement`. Chapters 3 and 4 both highlighted the *Parallele* substitution as the most frequent substitution to feature as an explanation for harmonic disagreement (this being equivalent to the relative key relationship). This prevalence was further confirmed in Chapter 4 by Participant 2 in the interview section of the study.

In contrast to Chapters 3 and 4, Chapter 5 found that the *Leittonswechsel* was the most frequent substitution capable of explaining harmonic disagreement, though the *Parallele* substitution was the second most common explanation of harmonic disagreement. Chapter 6 found that the *Parallele* substitution was the most frequently suitable explanation of harmonic disagreement among music analysts, followed by the *Leittonswechsel*. Interestingly, the participants in Chapter 3 were professional transcribers with musical training/education. Therefore, it might be this musical training/education that led to agreement between the annotators from Chapter 3 and the analysts. The *Variante* substitution was most frequently suitable for explaining the disagreement between popular musicians and composers in Chapter 6. On the other hand, we could categorise classical musicians and analysts as score-focused musicians who work from a text that is considered authoritative, and popular musicians and composers as musicians that either create their own music, or adapt the music of others. This difference — between those that work with music from an analytical and score-focused approach versus those that relate to the music through a more adaptive and creative manner, could be the cause of the differences in harmonic annotation.

Across all of the studies in this thesis, the *Parallele* substitution features prominently in explaining annotator disagreement. This finding concurs with the work of Carol Krumhansl, Jamshed Bharucha and Edward Kessler,[3] who found that participants perceive a key and its relative minor as most closely related.

3. Krumhansl, Bharucha, and Kessler, "Perceived Harmonic Structure of Chords in Three Related Musical Keys."

7.2.3 The most common sites of harmonic disagreement were at points of harmonic change

The findings of Chapters 3–5 identified points of harmonic changes as a potential cause of inter-annotator disagreement. In this type of disagreement, the annotators often agreed on the chords involved in a harmonic change, but disagreed on where precisely the change occurred. In Chapter 4's interview study, participants discussed having to correct the position of chords as a reason for chord edits. This cause of disagreement was particularly prominent when the harmony and vocal parts did not align; for example, in 'Little Bit O' Soul' (Chapter 6) at the point where the voice begins before the harmony changes (bars 4–5). Notably, in Chapter 4, 63% of the harmonic disagreement in 'Over the Rainbow' by Israel 'Iz' Kamakawiwo'ole was at points of harmonic change. This cause of annotator disagreement was particularly prominent in Chapter 6 where 39% of the disagreement arose at points of harmonic change, making it the most popular prediction for harmonic disagreement (20% higher than the next most common: disagreement recurring with a prominent musical feature).

7.3 Implications and Recommendations for future research

My thesis provides methodological novelty within both music theory and MIR. I have shown how the field of MIR, and specifically the sub-domain of Automatic Chord Estimation (ACE), can incorporate music theory to overcome some of the current limitations in extracting high level musical features. Music theory is often overlooked within MIR due to the requirement of a score to perform analyses, however my research shows the application of music theory beyond the study of scores, enabling notable computational advances through its utilisation. Within the field of music theory itself, my work also provides methodological novelty, not only by confirming the work of researchers such as Nicole Biamonte (2010), Walter Everett (2004) and Christopher Doll (2007) in Riemann's relevance across varying genres of music,[4] but also in providing a potential 'peace offering' in the current polemic between score and non-score aural literacy in academia.[5] My findings from Chapter 5 highlight the potential for music theory to not require a score, in-turn suggesting that score based aural literacy may not be essential as we can adapt music theoretical approaches to a non-score based curriculum.

4. Biamonte, "Triadic Modal and Pentatonic Patterns in Rock Music"; Everett, "Making Sense of Rock's Tonal Systems"; Doll, *Listening to Rock Harmony*.

5. Charlotte C Gill, *This article is more than 3 years old Music education is now only for the white and the wealthy*, Available at: https://www.theguardian.com/commentisfree/2017/mar/27/music-lessons-children-white-wealthy, March 2017; Ian Pace, *This romanticisation of musical illiteracy is risky*, Available at: https://www.theguardian.com/education/2017/apr/05/this-romanticisation-of-musical-illiteracy-is-risky, April 2017; Jon Henschen, *The tragic decline of music literacy and quality*, Available at: https://www.intellectualtakeout.org/article/tragic-decline-music-literacy-and-quality/, August 2018; Andrew Mellor, *Academics who dismiss musical literacy have confused recreation with study*, Available at: https://www.rhinegold.co.uk/classical_music/academics-dismiss-musical-literacy-confused-recreation-study/, October 2018.

For the most prominent similarity application — music recommendation — being able to incorporate important high-level music features, and further understand the relationship between pieces harmonically, could both improve current recommendation algorithms and provide new ways for determining similarity. Cross-genre recommendations that explore the complexities of music similarity could enable us to expand our musical horizons, and focus on different features of musical similarity other than genre or mood. The ability to reveal the qualities in two pieces of music that makes them likeable could present significant commercial advantage; for example, to predict, and therefore recommend, that I would love both Vaughan Williams 'Lark Ascending' and the Plain White T's 'Hey There Delilah'. These pieces are not similar in terms of their mood (Williams' piece features an intricate progression of moods, and the Plain White T's' song reflects both happiness and loss), or their genre (one is classical, the other emo-pop). This thesis has explored one music-theoretical approach and how it relates to similarity, yet there are still many more music-theoretical approaches to be explored, which have potential to further improve how we recommend music.

By adding to our knowledge of harmonic perception, this thesis has shown that it is possible to see chords related by harmonic substitutions as 'similar'. Further perception studies should explore whether participants see chords related by harmonic substitutions as more similar to each other than chords that are not related. This new knowledge of harmonic perception is also applicable to music plagiarism, where one could adopt a theoretical approach that observes the similarity between chords as critical in whether the song sounds similar to the human ear. My research confirms the recent discussions of Charles Cronin (2018), who stated that harmony was one of the most central parts of what makes a piece of music (including melody and rhythm as well), and that these should be the focus of music plagiarism investigations.[6] I have already explored the use of Riemannian theory in one copyright infringement case — Martin Harrington and Thomas Leonard vs. Ed Sheeran. If more music theoretical approaches align with audible music similarity, applying these to explain the perceptual similarities between pieces will enable us to create aligned methodologies for determining substantial similarities between pieces of music that are subject to copyright cases.

Not only does this research have applications in determining the similarity of pieces based on their harmony, but it could also have applications in helping musicians find simpler harmonic progressions to play. As discussed by Participant 2 in Chapter 4, harmonic substitutions can enable us to seek alternative chords to play for a complex passage (chords that sounds similar to the original transcription). Tools such as Chordify could use this to enable 'alternative' chord progressions for beginners who may find certain chords too demanding to play — meaning it could also be a useful didactic tool.

6. Cronin, "Seeing Is Believing: The Ongoing Significance of Symbolic Representations of Musical Works in Copyright Infringement Disputes."

This thesis only had scope to explore one music theoretical approach's application to music similarity: Riemannian theory. However, using the score (in Chapters 3, 4, 6) enabled me to utilise some aspects of Schenkerian analysis, such as prolongation and granularity. Previous researchers have posited that these aspects of similarity are important to audible similarity, such as Steve Larson (1997) and Alan Cadwallader and William Pastille (1992).[7] Future research that enables the computational implementation of these aspects of Schenkerian analysis could further aid our understanding of disagreement. The research of Alan Marsden and his collaborators have paved the way for utilising these approaches computationally.[8] As discussed in Chapter 1, each computational Schenkerian approach often features specific aspects of Schenkerian analysis that are deemed relevant. I suggest we look only at prolongation and granularity and adapt these current computational applications to enable this. Therefore, further work in optical music recognition and music encoding methods could aid in our ability to utilise these theories and also other music theories (not able to be explored in this thesis) in understanding music perception and music recommendation.

Research into ACE algorithms has extensively explored harmonic disagreement. As discussed in Chapter 3, where the study used a dataset previously collected for incorporating annotator disagreement in ACE algorithms,[9] current metrics used to measure annotator disagreement commonly focus on pitch-class agreement. These ACE algorithms currently paint too bleak a picture of the agreement between annotators. As discussed above, Vincent Koops et al. (2019) detailed a 69% agreement between the annotators using the common MIREX evaluation measures that determine the similarity of chords looking at the root-note, or the bass note and its third. Whereas, using a harmonic function analysis, we can explain a further 16% of disagreement in this dataset (totalling 82% using in this thesis's methodology compared to the 69% detailed by H. Vincent Koops et al. (2019)).[10] Therefore, Riemannian theory could and should be used to improve ACE algorithms' ability to extract chords and reflect human harmonic perception. Future research should investigate adding harmonic function to the current MIREX evaluation metrics for harmonic similarity, to enable us to show harmonic similarity and varying levels of musical understanding.

As this thesis has shown, the relationship between chords in terms of their substitution is what is essential to understanding annotator disagreement. Computational analysis of harmonic similarity could use this relationship to establish related chords by defining three rules based on these functions (for example in a natural language) to allow us to utilise this relationship. The MIR community has acknowledged the importance of harmony, but previously has not utilised music-theoretical approaches due to the lack

7. Larson, "The Problem of Prolongation in 'Tonal' Music: Terminology, Perception, and Expressive Meaning."; Cadwallader and Pastille, "Schenker's High-level Motives."

8. Marsden, Hirata, and Tojo, "Towards Computable Procedures for Deriving Tree Structures in Music: Context Dependency in GTTM and Schenkerian Theory."

9. Koops et al., "Annotator Subjectivity in Harmony Annotations of Popular Music."

10. Ibid.

of computer-readable scores.[11] The work of Jose Pedro Magalhaes and W. Bas de Hass (2011), and Tsuing-Ping Chen and Lin Su (2018) combined, would enable both the extraction of chord labels for music and then the determination of its functional element. Combining these methods computationally would enable us to utilise this harmonic relationship in a variety of MIR tasks, including recommendation, feature extraction, music structure analysis, and emotion analysis and classification. We could also look to utilise the vast range of chord annotations and guitar tab available online. These existing corpora could provide a substantial start to this process as they provide an existing collection of chord labelled music, requiring only the determination of its function.

Riemannian theory can therefore be used to improve machine learning and artificial intelligence (AI) approaches to music similarity. This could include enriching annotation training data as a basis for better optimised AI jobs. Importantly, how can we aim to replace human annotations if as humans we disagree on aural harmony? Current ground truth datasets utilise a single set of 'truth' between annotators. CASD was created to highlight this annotator disagreement, yet still MIR aims to use a single ground truth. Arguably, my thesis shows that this 'truth' is non existent. As a music theorist I could argue that the ground truth is an iterative and collaborative process, finding one 'true' analysis takes both time and collaboration between analysts. Therefore, is it right that the aim of MIR is to find the perfect/correct analysis or answer to a music theoretical question, with the constant growth to meet 100% accuracy? Yet, can ACE really hit 100% chord extraction accuracy if annotators cannot agree 100% of the time with each other on aural chord annotation? Therefore, should we re-shape the aim of MIR? Instead, should we aim to replicate human harmonic extraction, or provide a set of all possible annotations to speed up this collaborative iterative process of finding the one 'true' analysis? By working together, AI could enable music theorists to explore more pieces of music, create analyses across entire corpora and composers' vast works. But importantly, even when using computational methods we must not forget that we are not always going to be able to find a 'perfect' answer, as this 'true' or 'perfect' answer may not exist in the first place. Arguably, we could see that there is an ultimate glass ceiling, a point at which we may not want to, or be able to, improve these algorithms any further.

Breaking through perceived glass ceilings is a metaphor not just of my research's ability to resolve the current limitations of MIR, but also extensively represents my academic career to date. For both the discipline of MIR and music theory, there has been an apparent resistance and a lack of understanding in how they can aid one another. I represent a minority of academics whose research falls within both. Though received well, my conference papers often fall within the 'other' category in a program — where those papers whose work falls outside the fields usual scope. I hope my research has highlighted for both fields the importance of cross-discipline research; work that draws

11. Knees and Schedl, "A Survey of Music Similarity and Recommendation from Music Context Data."

upon the strength, knowledge and skills of researchers in both fields. Analogously, I wish to reflect on the glass ceilings I have also broken through being female within music analysis, MIR and also within the fields of music and engineering. Neither discipline has a balance or prominence of women; both are highly male dominated. There is an interesting parallel to be drawn between my advocacy for MIR and music theory to collaborate and promote joint scholarship, and the glass ceilings I shatter as a women within these fields.

Appendix A

Screenshots of 'Music Similarity Study 1'

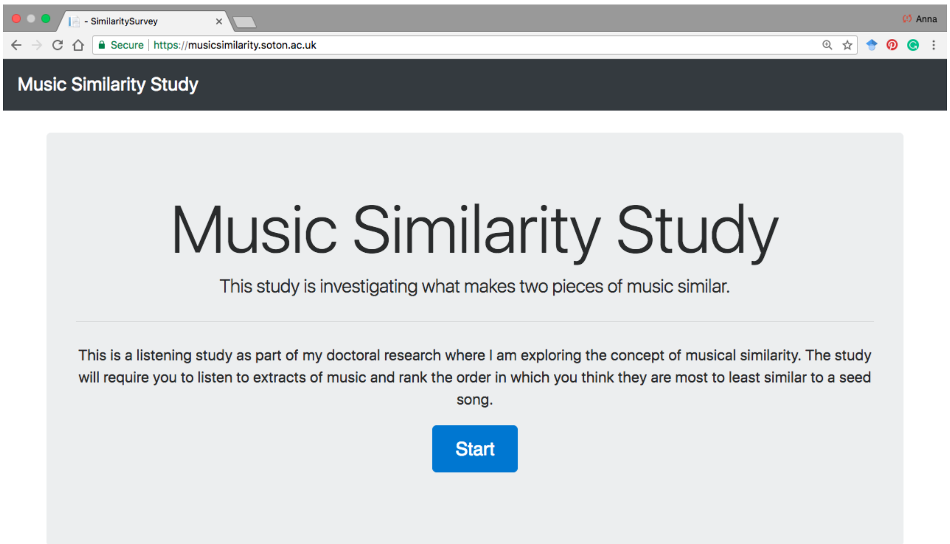


FIGURE A.1: The first page of 'music similarity study 1': the welcome page with a brief explanation of what the study entails.

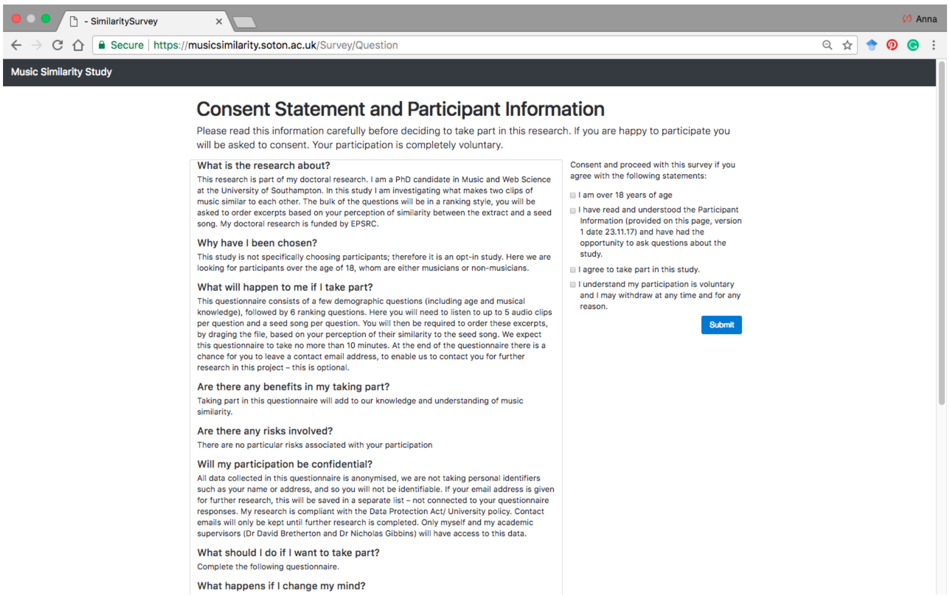


Figure A.2: The consent and information page of 'music similarity study 1'. This page was important for staying within ethics guidelines.

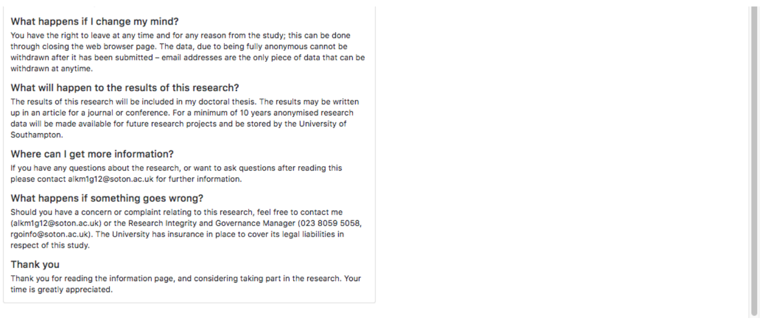


Figure A.3: Consent page cont.

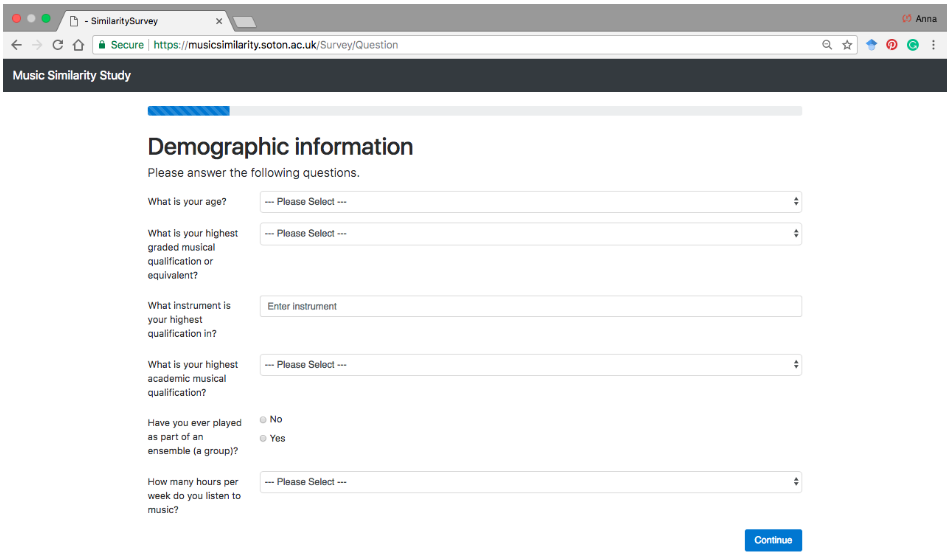


FIGURE A.4: The demographic page of 'music similarity study 1', where information such as age category, instrument qualifications, listening hours and academic music qualifications were collected.

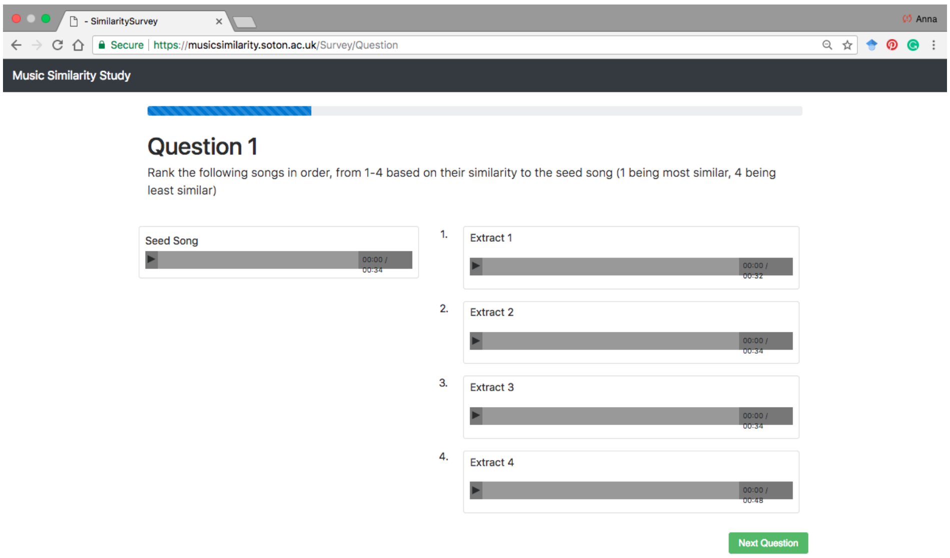


FIGURE A.5: Question 1 of 'music similarity study 1', with the extracts chosen based on form.

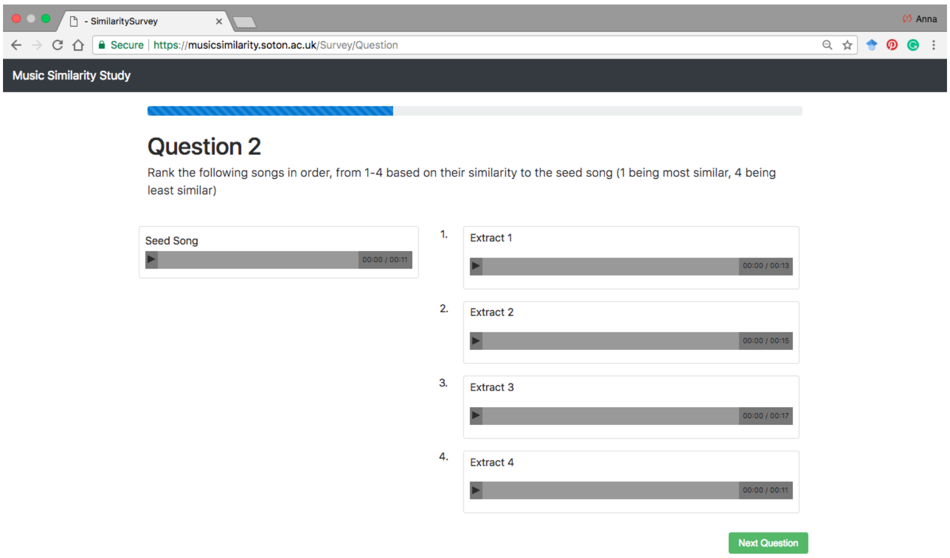


FIGURE A.6: Question 2 of 'music similarity study 1', with the extracts chosen based on tempo.

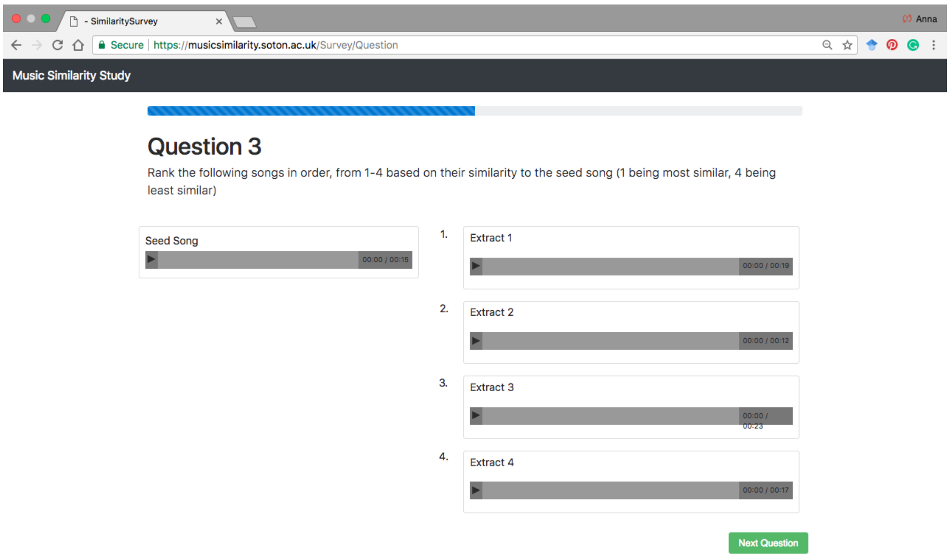


FIGURE A.7: Question 3 of 'music similarity study 1', with the extracts chosen were chosen according to melodic line.

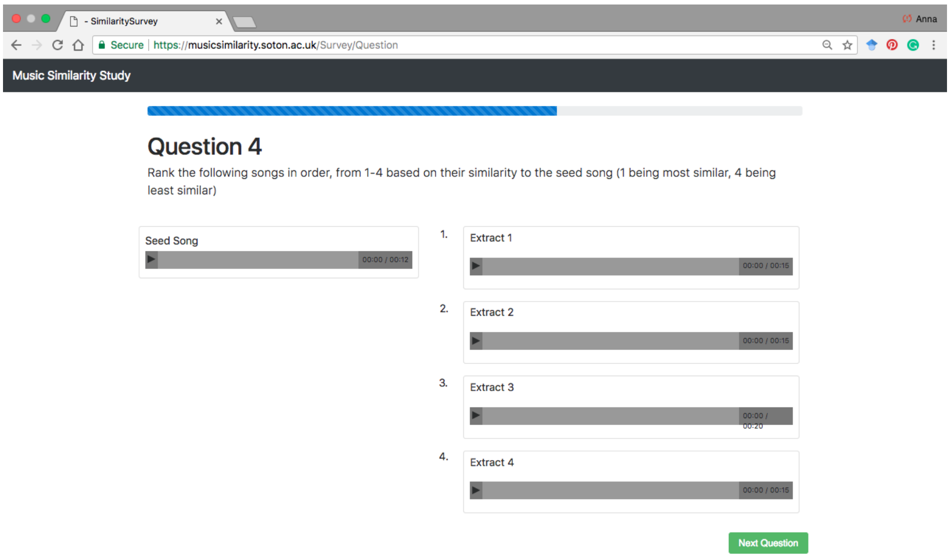


FIGURE A.8: Question 4 of 'music similarity study 1', where the extracts were chosen according to aspects of Schenkerian voice-leading.

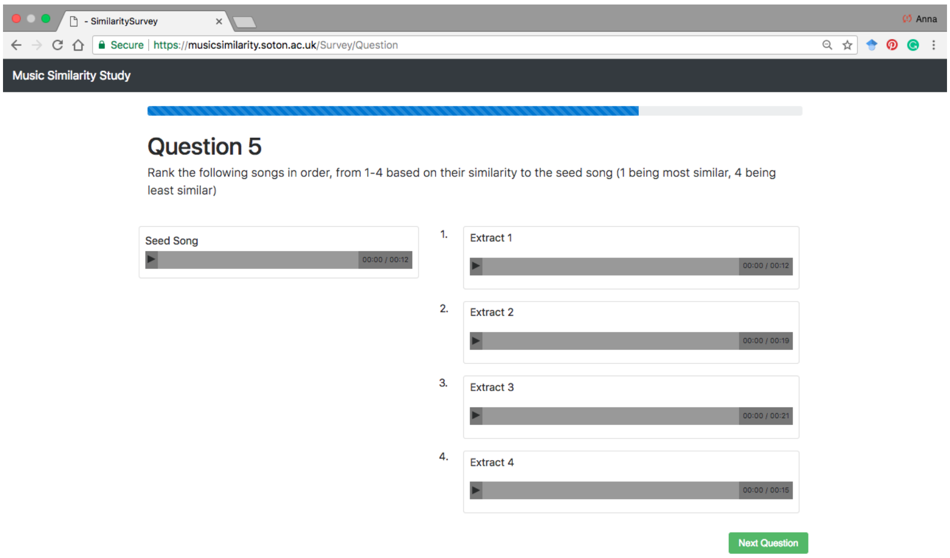


FIGURE A.9: Question 5 of 'music similarity study 1', where the extracts were chosen according to harmony.

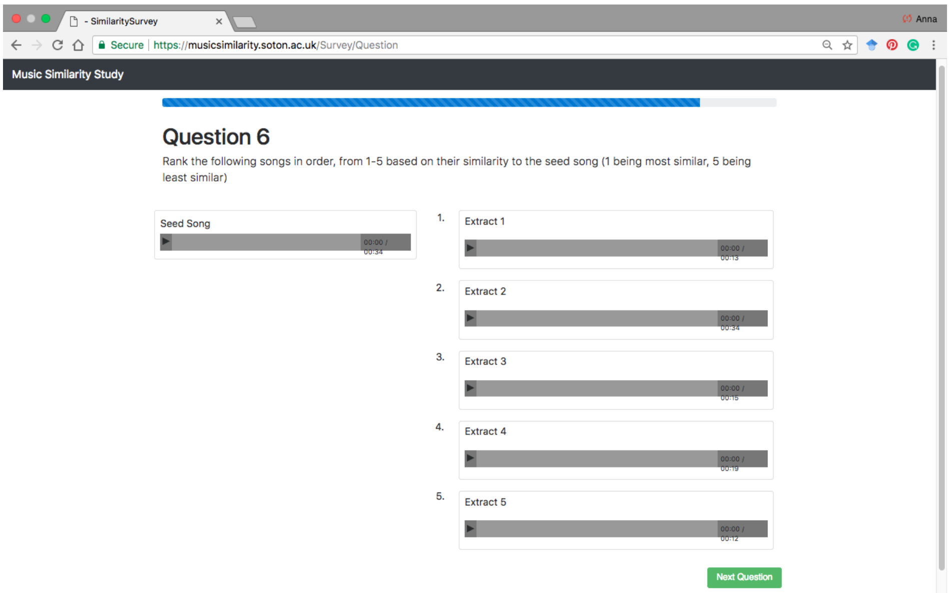


FIGURE A.10: Question 6 of 'music similarity study 1', where the extracts chosen were those ranked in position 1 for Questions 1–5.

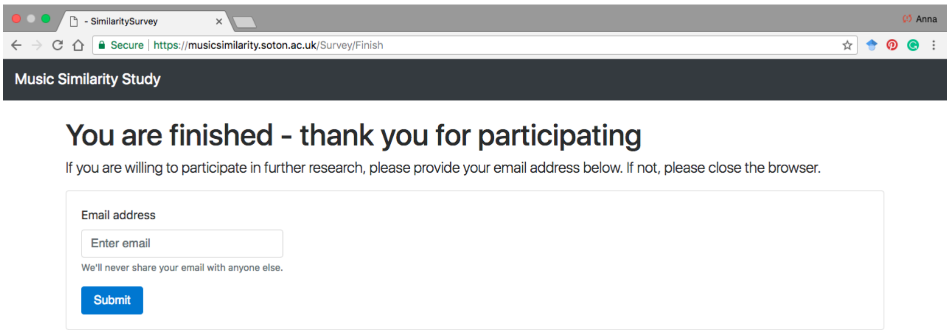


FIGURE A.11: The completion page of 'music similarity study 1', a participant could leave an email here if they were happy to be contacted for further research.

Bibliography

Agmon, Eytan. "Functional Harmony Revisited: A Prototype-Theoretic Approach." *Music Theory Spectrum* 17, no. 2 (1995): 196–214.

Ahlbäck, Sven. "Melodic Similarity as a Determinant of Melody Structure." *Musicae Scientiae* Discussion Forum 4A (2007): 253–280.

Aucouturier, Jean-Julien, Jean-Julien Aucouturier, François Pachet, and François Pachet. "Music similarity measures: What's the use." In *Proceedings of the 3rd International Society for Music Information Retrieval (ISMIR) Conference*, 157–163. Paris, 2002. ISBN: 2844261663.

Bello, Juan Pablo, Giuliano Monti, and Marnk B. Sandler. "Techniques for Automatic Music Transcription." In *Proceedings of the 11th International Society of Music Information Retrieval Conference* (Massachusetts), 2000, 23–25.

Bello, Juan Pablo, and Jeremy Pickens. "A Robust Mid-Level Representation for Harmonic Content in Music Signals." In *Proceedings of the 6th International Society for Music Information Retrieval (ISMIR) Conference* (London), 2005, 304–311.

Bent, Ian D., and Anthony Pople. "Analysis." In *Grove Music Online*, edited by Stanley Sadie. Oxford University Press, 2001.

Berenzweig, Adam, Beth Logan, Daniel P.W. Ellis, and Brian Whitman. "A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures." *Computer Music Journal* 28, no. 2 (2004): 63–76.

Bernstein, David W. "Nineteenth Century Harmonic Theory." In *The Cambridge History of Western Music Theory*, 3rd, edited by Thomas Christensen, 778–817. Cambridge: Cambridge University Press, 2006.

Biamonte, Nicole. "Modal Function in Rock and Heavy Metal Music." In *L'analyse musicale aujourd'hui*, edited by Modher. Ayari, Jean-Michel Bardex, and Xavier Hascher, 275–290. Université: Delatour France, 2012.

———. "Triadic Modal and Pentatonic Patterns in Rock Music." *Music Theory Spectrum* 32, no. 2 (2010): 95–110.

Bigo, Louis, Moreno Andreatta, Jean-Louis Giavitto, Oliver Michel, and Antoine Spicher. "Computation and Visualisation of Musical Structures in Chord-Based Simplicial Complexes." In *Proceedings of the International Conference on Mathematics and Computation in Music* (Berlin), 2013, 38–51.

Bogdanov, Dmitry, Joan Serra, Nicolas Wack, Perfecto Herrera, and Xavier Serra. "Unifying low-level and high-level music similarity measures." *IEEE Transactions on Multimedia* 13, no. 4 (2011): 687–701.

Bradlow, Eric T., and Peter S. Fader. "A Bayesian Lifetime Model for the 'Hot 100' Billboard songs." *Journal of the American Statistical Association* 96, no. 454 (2001): 368–381.

Bragg, Jonathan, Elaine Chew, and Stuart Shieber. "Neo-Riemannian Cycle Detection with Weighted Finite-State Transducers." In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 399–404.

Breese, John, David Heckerman, and Carl Kadie. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering." In *Proceedings of the ACM Conference on Computer Human Interaction* (Denver), 1998, 210–217.

Bresson, Jean, Agon Carols, and Gérard Assayag. "OpenMusic: Visual Programming Environment for Music Composition, Analysis and Research." In *Proceedings of the ACM MultiMedia (OpenSource Software Competition),* (Arizona), 2011.

Brown, Matthew. *Explaining Tonality: Schenkerian Theory and Beyond*. New York: University of Rochester Press, 2005.

Burgoyne, John Ashley. "Stochastic Processes and Database-Driven Musicology." PhD diss., 2012.

Burgoyne, John Ashley, J. Wild, and I. Fujinaga. "An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis." In In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference*, 633–638. Miami, 2011.

Byrd, Donald, and Eric Isacson. "A Music Representation Requirement Specification for Academic." *Computer Music Journal* 27, no. 4 (2003): 43–57.

C Gill, Charlotte. *This article is more than 3 years old Music education is now only for the white and the wealthy*. Available at: https://www.theguardian.com/commentisfree/2017/mar/27/music-lessons-children-white-wealthy, March 2017.

Cabral, Giordano, François Pachet, Jean-Pierre Briot, and S Paris. "Automatic X Traditional Descriptor Extraction: The Case of Chord Recognition." In *Proceedings of the 6th International Society of Music Information Retrieval (ISMIR) Conference* (London), 2005, 444–449.

Cadwallader, Allen, and David Gagne. *Analysis of Tonal Music: A Schenkerian Approach.* New York: Oxford University Press, 2011.

Cadwallader, Allen, and William Pastille. "Schenker's High-level Motives." *Journal of Music Theory* 36, no. 1 (1992): 119–148.

Camboropoulos, Emilios. "Musical Parallelism and Melodic Segmentation: A Computational Approach." *Music Perception: An Interdisciplinary Journal* 23, no. 3 (2006): 249–268.

Caplin, William E. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart and Beethoven.* New York: Oxford University Press, 1998.

———. "What are Formal Functions?" In *Music Form, Forms, Formenlehre: Three Methodological Reflections*, 2nd, edited by Pieter Bergé, 21–40. Leuven: Leuven University Press, 2010.

Capuzzo, Guy. "Neo-Riemannian Theory and the Analysis of Pop-Rock Music." *Music Theory Spectrum* 26, no. 2 (2004): 177–199.

Casey, Michael. "General Sound Classification and Similarity in MPEG-7." *Organised Sound* 6, no. 2 (2001): 153–164.

Casey, Michael .A., Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcom Slaney. "Content-Based Music Information Retrieval: Current Directions and Future Challenges," 96, no. 4 (2008): 668–696.

Chai, Wei. "Semantic Segmentation and Summarization of Music: Methods based on Tonality and Recurrent Structure." *IEEE Signal Processing Magazine* 23, no. 2 (2006): 124–132.

Chen, Ruofeng, Weibin Shen, Ajay Srinivasamurthy, and Parag Chordia. "Chord Recognition using Duration-Explicit Hidden Markov Models." In *Proceedings of the 13th International Society for Music Information Retrieval (ISMIR) Conference* (Portugal), 2012, 445–450.

Chen, Tsung-Ping, and Li Su. "Functional Harmony Recognition of Symbolic Music Data with Multi-Task Recurrent Neural Networks." In *Proceedings of the 19th International Society for Music Information Retrieval (ISMIR) Conference* (Paris), 2018, 90–97.

Clark, Suzannah. "On the Imagination of Tone in Schubert's Liedesend (D473), Trost (D523), and Gretchens Bitte (D564)." In *The Oxford Handbook of neo-Riemannian Music Theories*, edited by Edward Gollin and Alexander Rehding, 294–321. New York: Oxford University Press, 2011.

Clercq, Trevor de, and David Temperley. "A corpus analysis of rock harmony." *Popular Music* 30, no. 1 (2011): 47–70.

Cohn, Richard. "As Wonderful as Star Clusters: Instruments for Gazing at Tonality in Schubert." *19th Century Music* 22, no. 3 (1999): 213–232.

———. *Audacious Euphony: Chromaticism and the Triad's Second Nature.* New York: Oxford University Press., 2012.

———. "Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective." *Journal of Music Theory* 42, no. 2 (1998): 167–180.

———. "Maximally Smooth Cycles, Hexatonic Systems, and the Analysis of Late-Romantic Triadic Progressions." *Music Analysis* 15, no. 1 (1996): 9–40.

———. "Neo-Riemannian Operations, Parsimonious Trichords and their "Tonnetz" representations." *Journal of Music Theory* 41, no. 1 (1997): 1–66.

Cohn, Richard, Brian Hyer, Carl Dahlhaus, Julian Anderson, and Charles Wilson. *Harmony.* Available at https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000050818, 2001.

Coker, Jerry. *Complete Method for Improvisation: For All Instruments.* Essex: Alfred Publishing (UK) Ltd, 1997.

Cook, Nicholas. *A Guide to Musical Analysis.* London: Butler / Tanner Ltd., 1987.

———. "Epistemologies of Music Theory." In *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 78–105. Cambridge: Cambridge University Press, 2002.

———. "Musical Form and the Listener." *The Journal of Aesthetic and Art Criticism* 44, no. 1 (1987): 23–29.

Cook, Nick. "Towards the Compleat Musicologist?" Keynote address at the 6th International Society for Music Information Retrieval (ISMIR) Conference, London, 2005.

Council on Library and Information Resources. "Structural Analysis of Large Amounts of Music Information (SALAMI)." CLIR. 2004. Accessed April 28, 2018. https://www.clir.org/pubs/reports/pub151/case-studies/salami.

Cronin, Charles. "Seeing Is Believing: The Ongoing Significance of Symbolic Representations of Musical Works in Copyright Infringement Disputes." *Colorado Technology Law Journal* 16, no. 2 (2018): 225–248.

Crowder, Robert G. "Perception of the Major/Minor Distinction: I. Historical and Theoretical Foundations." *Psychomusicology: Music, Mind, and Brain* 4, no. 1 (1984): 3–12.

Dahlhaus, Carl, Julian Anderson, Charles Wilson, Richard Cohn, and Brian Hyer. "Harmony." In *Oxford Music Online*. Oxford University Press, 2009.

Daintith, John, and Edmund Wright. *Code*. Available at: http://www.oxfordreference.com/view/10.1093/acref/9780199234004.001.0001/acref-9780199234004, 2008.

Davies, Matthew, and Sebastian Böck. "Evaluating the Evaluation Measures for Beat Tracking." In *Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference* (Taipei), 2014, 637–642.

Davies, Stephen. "Transcription, Authenticity and Performance." *British Journal of Aesthetics* 18, no. 1 (1988): 216–227.

De Roure, David, J. Stephen Downie, and Ichiro Fujinaga. "SALAMI: Structural Analysis of Large Amounts of Music Information." In *Proceedings of the 10th UK e-Science All Hands Meeting* (York), 2011. https://eprints.soton.ac.uk/271171/.

Deliége, Irene. "Similarity Relations in Listening to Music: How do they come into Play?" *Musicae Scientiae* Discussion Forum 4A (2007): 9–37.

Devaney, Johanna, Claire Arthur, Nathaniel Condit-Schultz, and Kristen Nisula. "Theme and Variation Encodings with Roman Numerals (TAVERN): A New Data Set for Symbolic Music Analysis." In *Proceedings of the 16th annual International Society for Music Information Retrieval (ISMIR) Conference* (Málaga), 2015, 728–734.

Dewi, Kadek Cahya, Luh Arida, and Ayu Rahning. "Music Recommendation Based on Audio Similarity Using K-Nearest Neighbor." In *Proceedings of the 1st ACIKITA International Conference of Science and Technology (AICST)*, 124–132. 978. 2011.

Doan, Anhai, Raghu Ramakrishnan, and Alon Y. Halevy. "Crowdsourcing systems on the world-wide web." *Communications of the ACM* 54, no. 4 (2011): 86–96.

Doll, Christopher. *Hearing Harmony: Toward a tonal theory for the Rock Era*. Michigan: University of Michigan Press, 2017.

———. *Listening to Rock Harmony*. Columbia: Columbia University, 2007.

Douthett, Jack, and Peter Steinbach. “Parsimonious Graphs: A Study in Parsimony, Contextual Transformations, and Modes of Limited Transpositions.” *Journal of Music Theory* 42, no. 2 (1998): 241–263.

Downie, J. Stephen. “The Music Information Retrieval Evaluation Exchange (2005-2007): A Window into Music Information Retrieval Research.” *Acoustical Science and Technology* 29, no. 4 (2008): 247–255.

Drabkin, William. “Prolongation.” In *Grove Music Online*, edited by Stanley Sadie. Oxford University Press, 2001. http://www.oxfordmusiconline.com/subscriber/article/grove/music/22408.

Duvall, Paul H, and Mark L Block. *Complaint for Copyright Infringement, HaloSongs, Inc., Martin Harrington, and Thomas Leonard vs. Edward Christopher Sheeran*, 2016.

Eitan, Zohar, and Roni Y. Granot. “Primary versus Secondary Musical Parameters and the Classification of Melodic Motives.” *Musicae Scientiae* Discussion Forum 48 (2009): 139–179.

Engebretsen, Nora. “Neo-Riemannian Perspectives on the Harmonieschritte, with a Translation of Riemann’s Systematik der Harmonieschritte.” In *The Oxford Handbook of neo-Riemannian Music Theories*, edited by Edward Gollin and Alexander Rehding, 351–381. New York: Oxford University Press, 2011.

Everett, Walter. “Making Sense of Rock’s Tonal Systems.” *Music Theory Online* 10, no. 4 (2004).

Fink, Arlene. *How to Ask Survey Questions*. 2nd. California: Sage Publications, Inc., 2003.

Foote, Jonathan. “Automatic Audio Segmentation using a Measure of Audio Novelty.” In *Proceedings of the IEEE International Conference on Multimedia and Expo* (New York), 2000, 452–455.

Forte, Allen, and Steven E. Gilbert. *Introduction to Schenkerian Analysis*. New York: W. W. Norton & Company, Inc., 1982.

Fruehwald, E. Scott. “Copyright Infringement of Musical Compositions: A Systematic Appproach.” *Akron Law Review* 26, no. 1 (1993): 15–44.

Fujishima, Takuya. “Realtime Chord Recognition of Musical Sound: A System using Common Lisp Music.” In *Proceedings of the International Computer Music Conference* (Beijing), 1999, 464–467.

Fujita, Tetsuya, Yuji Hagino, Hajime Kubo, and Goro Sato. *The Beatles: complete scores.* 1136. Hal Leonard Publishing, 1993.

Gennet, Robbie. *Why is Sheet Music Still Considered Necessary for Music Education?* Available at: https://www.huffpost.com/entry/why-is-sheet-music-still_b_7975400, December 2016.

Giraud, Jennifer, Richard Groult, and Levé Florence. "Computational Analysis of Musical Form." In *Computational Music Analysis*, edited by David Meredith, 113–136. New York: Springer, 2016.

Goldman, Andrew, Tyreek Jackson, and Paul Sajda. "Improvisation Experience Predicts how Musicians Categorize Musical Structures." *Psychology of Music*, 2018, 1–17.

Gollin, Edward. "Some Aspects of Three-Dimensional 'Tonnetz'." *Journal of Music Theory* 42, no. 2 (1998): 195–206.

Gomes, Carlos, Daniel Schneider, Katia Moraes, and Jano De Souza. "Crowdsourcing for Music: Survey and Taxonomy." In *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, 832–839. IEEE, 2012.

Goto, Masataka. "A Chorus-Section Detecting Method for Musical Audio Signals." In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Hong Kong), 2003, 437–440.

Haas, W. Bass . de, José Pedro Magalhães, and Frans Wiering. "Improving audio chord transcription by exploiting harmonic and metric knowledge." In *Proceedings of the 13th International Society of Music Information Retrieval (ISMIR) Conference* (Portugal), 2012, 295–300.

Harrison, Daniel. *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of its Precedents*. Chicago: The University of Chicago Press, 1994.

Hawkins, Stan. "Musicological Quagmires in Popular Music." *Popular Musicology Online*, 2011.

Henschen, Jon. *The tragic decline of music literacy and quality*. Available at: https://www.intellectualtakeout.org/article/tragic-decline-music-literacy-and-quality/, August 2018.

Hepokoski, James. "Sonata Theory and Dialogic Form." In *Music Form, Forms, Formenlehre: Three Methodological Reflections*, 2nd, edited by Pieter Bergé, 41–46. Leuven: Leuven University Press, 2010.

Hickey, Andrew. *The Beatles In Mono*. United States: Lulu Press, 2010.

Hoffman, Justin. *Listening with Two Ears: Conflicting Perceptions of Space in Tonal Music*. Columbia: Columbia University, 2011.

Howe, Jeff. "The Rise of Crowdsourcing." *Wired magazine* 14, no. 6 (2006): 1–4.

Hu, Xiao, and Stephen J. Downie. In *Proceedings of the 11th International Society for Music Information Retrieval (ISMIR) Conference* (Utrecht), 2010, 619–624.

Huron, David. "A Psychological Approach to Musical Form: The Habituation-Fluency Theory of Repetition." *Current Musicology* 96, no. 1 (2013): 7–35.

Hyer, Brian. "Tonality." In *The Cambridge History of Western Music Theory*, 3rd, edited by Thomas Christensen, 726–752. Cambridge: Cambridge University Press, 2006.

———. "What is a Function?" In *The Oxford Handbook of neo-Riemannian Music Theories*, edited by Edward Gollin and Alexander Rehding, 92–139. New York: Oxford University Press, 2011.

Jairazbhoy, Nazir A. "The 'Objective' and Subjective View in Music Transcription." *Ethnomusicology* 21, no. 2 (1977): 263–273.

Janssen, Berit, Peter van Kranenburg, and Anja Volk. "Finding occurrences of melodic segments in folk songs employing symbolic similarity measures." *Journal of New Music Research* 46, no. 2 (2017): 118–134.

Kann, Hans. 'Preface' to *Variationen über ein Thema von Paganini Opus 35*. G. Heile Verlag, 1985.

Kassler, Michael. "A Trinity of Essays: Toward a Theory that is the Twelve-Note Class System; Toward Development of a Constructive Tonality Theory Based on Writings by Heinrich Schenker; Toward a Single Programming Language for Musical Information Retrieval." PhD diss., 1967.

———. "APL Applied in Music Theory." In *Proceedings of the International Conference on APL* (Texas), 1987, 209–214.

———. "Explication of the Middleground of Schenker's Theory of Tonality." *Miscellanea Musicologica: Adelaide Studies in Musicology* 9 (1977): 72–81.

———. *Proving Musical Theorems I: The Middleground of Heinrich Schenker's Theory of Tonality.* The University of Sydney, 1975.

Keifer, Andrea. "Civil Minutes from Complaint for Infringement, Pharrell Williams et al. vs. Brdigeport Music, Inc., et al." *United States District Court Central District Of California* Case 2:13 (2014): 1–28.

Kennedy, Michael. "Thematic Material." In *The Oxford Dictionary of Music*, 2nd. Oxford University Press, 2006.

Kirlin, Phillip B. "A Data Set for Computational Studies of Schenkerian Analysis." In *Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference* (Taipei), 2014, 213–218.

———. "A Lesson in Analysis from Heinrich Schenker: The C Major Prelude from Bach's Well-Tempered Clavier, Book I." In *Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference* (Taipei), 2014, 213–218.

———. "A Probabilistic Model of Hierarchical Music Analysis." PhD diss., 2014.

———. "Global Properties of Expert and Algorithmic Hierarchical Music Analyses." In *Proceedings of the 17th International Society for Music Information Retrieval (ISMIR) Conference* (New York), 2016, 640–646.

———. "Using Harmonic and Melodic Analyses to Automate the Initial Stages of Schenkerian Analysis." In *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference* (Kobe), 2009, 423–428.

Kirlin, Phillip B., and David D. Jensen. "Probabilistic Modelling of Hierarchical Music Analysis." In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 393–398.

———. "Using Supervised Learning to Uncover Deep Musical Structure." In *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (Texas), 2015, 1770–1777.

Kirlin, Phillip B., and David L. Thomas. "Extending a Model of Monophonic Hierarchical Music Analysis to Homophony." In *Proceedings of the 16th International Society for Music Information Retrieval (ISMIR) Conference* (Málaga), 2015, 715–721.

Kirlin, Phillip B, and Paul E. Utgoff. "A Framework for Automated Schenkerian Analysis." In *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference* (Philadelphia), 2008, 363–368.

Klapuri, Anssi. "Introduction to Music Transcription." In *Signal Processing Methods for Music Transcription*, edited by Anssi Klapuri and Manuel Davy, 3–20. New York: Springer US, 2006.

Knees, Peter, and Markus Schedl. "A Survey of Music Similarity and Recommendation from Music Context Data." *ACM Transactions on Multimedia Computing, Communications, and Applications* 10, no. 1 (2013): 1–21.

Knyt, Erinn E. "'How I Compose': Ferruccio Busoni's Views about Invention, Quotation, and the Compositional Process." *The Journal of Musicology* 27, no. 2 (2010): 224–264.

Koops, Hendrik Vincent. "Computational Modelling of Variance in Musical Harmony." PhD diss., Utrecht University, 2019.

Koops, Hendrik Vincent, Bas de Haas, John Ashley Burgoyne, Jeroen Bransen, and Anja Volk. *Harmonic Subjectivity in Popular Music.* Technical report UU-CS-2017-018. Department of Information and Computing Sciences, Utrecht University, 2017.

Koops, Hendrik Vincent, W. Bas de de Haas, John Ashley Burgoyne, Jeroen Bransen, Anna Kent-Muller, and Anja Volk. "Annotator Subjectivity in Harmony Annotations of Popular Music." *Journal of New Music Research* 48, no. 3 (2019): 232–252.

Krumhansl, Carol L. "Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations." *Journal of Music Theory* 42, no. 2 (1998): 256–281.

———. "Tonal Hierarchies and Rare Intervals in Music Cognition." *Music Perception: An Interdisciplinary Journal* 7, no. 3 (1990): 309–324.

Krumhansl, Carol L., Jamshed J. Bharucha, and Edward J. Kessler. "Perceived Harmonic Structure of Chords in Three Related Musical Keys." *Journal of Experimental Psychology: Human Perception and Performance* 8, no. 1 (1982): 24–36.

Kullback, S., and R. Leibler. "On information and sufficiency." *The Annals of Mathematical Statistics* 22, no. 1 (1951): 79–86.

Larson, Steve. "The Problem of Prolongation in 'Tonal' Music: Terminology, Perception, and Expressive Meaning." *Journal of Music Theory* 41, no. 1 (1997): 101–136.

Laurier, Cyril, Mohamed Sordo, Joan Serra, and Perfecto Herrera. "Music Mood Representations from Social Tags." In *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference* (Utrecht), 2009, 381–386.

Lee, Jin Ha. "How Similar is too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation." In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 109–114.

Levy, Mark, Mark Sandler, and Michael Casey. "Extraction of High-Level Musical Structure from Audio Data and its Application to Thumbnail Generation." In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing* (France), 2006.

Lewin, David. *Generalized Musical Intervals and Transformations.* Oxford: Oxford University Press, 2010.

Littlewood, Julian. *The Variations of Johannes Brahms.* Plumbago Books, 2004.

Logan, Beth, and Ariel Salomon. "A Music Similarity Function based on Signal Analysis." *IEEE International Conference on Multimedia and Expo (ICME)* (Tokyo), 2001, 745–748.

Lu, Lie, Muyuan Wang, and Zhangm Hong-Jiang. "Repeating Pattern Discovery and Structure Analysis from Acoustic Music Data." In *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval* (New York), 2004, 275–282.

Lubben, Joseph. "Schenker the Progressive: Analytic Practice in 'Der Tonwille'." *Music Theory Spectrum* 15, no. 1 (1993): 59–75.

Lucero, Mario J. *Music Streaming Services Mishandle our Data—and our Culture is Paying for it.* Available at: https://qz.com/1773480/the-problem-with-how-the-music-streaming-industry-handles-data/, January 2020.

Luger, George F., and William A. Stubblefield. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving.* Benjamin-Cummings Publishing Company, 1992.

Maddage, Namunu Chinthaka. "Content-Based Music Structure Analysis." PhD diss., 2006.

Magalhães, Joseé, and W. Bas de Haas. "Functional Modelling of Musical Harmony: an Experience Report." *The ACM Special Interest Group on Programming Languages Notices* 46, no. 9 (2011): 156–162.

Marsden, Alan. "Automatic Derivation of Musical Structure: A Tool for Research on Schenkerian Analysis." In *Proceedings of the 8th International Society for Music Information Retrieval (ISMIR) Conference* (Vienna), 2007, 55–58.

———. "Generative Structural Representation of Tonal Music." *Journal of New Music Research* 34, no. 4 (2005): 409–428.

———. "Interrogating Melodic Similarity: A Definitive Phenomenon or the Product of Interpretation?" *Journal of New Music Research* 41, no. 4 (2012): 323–335.

———. "Representing Melodic Patterns as Networks of Elaborations." *Computers and Humanities* 35 (2001): 37–54.

———. "Schenkerian Analysis by Computer: A Proof of Concept." *Journal of New Music Research* 39, no. 3 (2010): 269–289.

———. "Software for Schenkerian Analysis." In *Proceedings of 37th the International Computer Music Conference* (Huddersfield), 2011, 673–676.

———. "Towards Schenkerian Analysis by Computer: A Reductional Matrix." In *Proceedings of 31st the International Computer Music Conference* (Barcelona), 2005, 247–250.

Marsden, Alan, Keiji Hirata, and Satoshi Tojo. "Towards Computable Procedures for Deriving Tree Structures in Music: Context Dependency in GTTM and Schenkerian Theory." In *Proceedings of the Sound and Music Computing (SMC) Conference* (Stockholm), 2013, 360–367.

Marsden, Alan, and Gerant A. Wiggins. "Schenkerian Reduction as Search." In *Proceedings of the Fourth Conference on Interdisciplinary Musicology* (Thessaloniki), 2008, 1–9.

Mauch, Matthias, Katy Noland, and Simon Dixon. "Using Musical Structure to Enhance Automatic Chord Transcription." In *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference* (Utrecht), 2009, 231–236.

Mavromatis, Panayotis, and Matthew Brown. "Parsing Context-Free Grammars for Music: A Computational Model of Schenkerian Analysis." In *Proceedings of the 8th International Conference on Music Perception and Cognition* (Illinois), 2004, 414–415.

McAdams, Stephen, and Daniel Matzkin. "Similarity, Invariance, and Musical Variation." *Annals of the New York Academy of Sciences* 930, no. 1 (2001): 62–76.

McVicar, Matt, Tim Freeman, and Tijl De Bie. "Mining the Correlation Between Lyrical and Audio Features and the Emergence of Mood." In In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference*, 783–788. Miami, 2011.

McVicar, Matt, Raúl Santos-Rodrıguez, Yizhao Ni, and Tijl De Bie. "Automatic Chord Estimation from Audio: A Review of the State of the Art." *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* 22, no. 2 (2014): 556–575.

Mellers, Wilfrid. *Twilight of the Gods; the Music of The Beatles.* 215. Richard Seaver Bks. New York: Viking Press, 1974.

Mellor, Andrew. *Academics who dismiss musical literacy have confused recreation with study.* Available at: https://www.rhinegold.co.uk/classical_music/academics-dismiss-musical-literacy-confused-recreation-study/, October 2018.

Meyer, Leonard. *Emotion and Meaning in Music.* Chicago: The University of Chicago Press, 1956.

———. *The Spheres of Music: A Gathering of Essays.* Chicago: The University of Chicago Press, 2000.

Milne, Andrew J., and Simon Holland. "Empirically Testing Tonnetz, Voice-Leading and Spectral Models of Perceived Triadic Distance." *Journal of Mathematics and Music: Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance* 10, no. 1 (2016): 59–85.

Mobasher, Bamshad, Honghua Dai, Tao Luo, and Miki Nakagawa. "Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data." In *Workshop Intelligent techniques for Web Personalization*, edited by Sarabjot Singh Anand and Bamshad Mobasher, 53–60. Berlin: S, 2001.

Müllensiefen, Daniel, and Kalus Frieler. "Modelling Experts' Notions of Melodic Similarity." *Musicae Scientiae* Discussion Forum 4A (2007): 183–210.

Müller, Meinard. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications.* Switzerland: Springer International Publishing, 2015.

Nawab, S. Hamid, Salma Ayyash, and Robert Wotiz. "Identification of Musical Chords using Constant–Q Spectra." In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP* (Massachusetts), 2001, 3373–3376.

Neal, Jason. *Defining Musical Similarity: Genre and Beyond.* http://www.asis.org/SocialMedia/?p=66. (Accessed: 25.06.2016). The Information Associaton for the Information Age (ASIS&T), 2014.

Neighbour, O.W. "Schoenberg [Schönberg], Arnold." In *Oxford Music Online.* Oxford University Press, 2001.

Ni, Yizhao, Matt McVicar, Raúl Santos-Rodrıguez, and Tijl De Bie. "An end-to-end Machine Learning System for Harmonic Analysis of Music." In *Proceedings of the IEEE Transactions on Audio, Speech and Language Processing* 20, no. 6 (2012): 1771–1783.

———. "Understanding Effects of Subjectivity in Measuring Chord Estimation Accuracy." *IEEE Transactions on Audio, Speech, and Language Processing* 21, no. 12 (2013): 2607–2615.

O'Grady, Terence J. *The Beatles, a Musical Evolution.* Music Series. United States: Twayne Publishers Inc., 1983.

Ockelford, Adam. "Implication and Expectation in Music: A Zygonic Model." *Psychology of Music* 31, no. 1 (2006): 81–142.

Oudre, Laurent, Yves Grenier, and Cédric Févotte. "Chord recognition using measures of fit, chord templates and filtering methods." In *Proceedings of the IEEE Workshop on Applied Signal Processing and Audio Acoustics*, 2012, 9–12.

Pace, Ian. *This romanticisation of musical illiteracy is risky.* Available at: https://www.theguardian.com/education/2017/apr/05/this-romanticisation-of-musical-illiteracy-is-risky, April 2017.

Palmer, Jason. "'Blurred Lines' Means Changing Focus: Juries Composed of Musical Artists Should Decide Music Copyright Infringement Cases, Not Lay Juries." *Vanderbilt Journal of Entertainment Technology Law* 18, no. 4 (2016): 907–934.

Paulus, Jouni, and Anssi Klapuri. "Music Structure Analysis by Finding Repeated Parts." In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia* (California), 2006, 59–68.

Pearsall, Edward R. "Harmonic Progressions and Prolongation in Post-Tonal Music." *Music Analysis* 10, no. 3 (1991): 345–355.

Pedler, Dominic. *The Songwriting Secrets of The Beatles.* London: Omnibus Press, 2003.

Peeters, Geoffroy, and Emmanuel Deruty. "Is Music Structure Annotation Multi-Dimensional? A Proposal for Robust Local Music Annotation." In *Proceedings of the 3rd Workshop on Learning the Semantics of Audio Signals* (Austria), 2009, 75–90.

Pinchot Kastner, Marianna, and Robert G. Crowder. "Perception of the Major/Minor Distinction: IV. Emotional Connotations in You." *Music Perception* 8, no. 2 (1990): 189–202.

Porter, Steven Clark. *Rhythm and Harmony in the Music of The Beatles.* 886. City University of New York, 1983.

Post, David. *Blurred Lines and Copyright Infringement.* Available at: https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/03/12/blurred-lines-and-copyright-infringement/, March 2015.

Raykar, Vikas C., and Shipeng Yu. "Eliminating Spammers and Ranking Annotators for Crowdsourced Labelling Tasks." *Journal of Machine Learning Research* 13 (2012): 491–518.

Regener, Eric. "Layered Music-Theoretic Systems." *Perspectives of New Music* 6, no. 1 (1996): 52–62.

Riedl, Christoph, Ivo Blohm, Jan Marco Leimeister, and Helmut Krcmar. "Rating Scales for Collective Intelligence in Innovation Communities: Why Quick and Easy Decision Making does not get it Right." In *Proceedings of 2010 International Conference on Information Systems* (St Louis), 2010.

Riemann, Hugo. *Harmony Simplified: or, the Theory of the Tonal Functions of Chords.* London: Augener, 1896.

———. "Ideas for a Study 'On the Imagination of Tone'." Translated by Robert W. Wason and Elizabeth West Marvin. *Journal of Music Theory* 36, no. 1 (1992): 81–117.

Rowley, Jennifer. "Conducting Research Interviews." *Management Research Review* 35, nos. 3/4 (2012): 260–271.

Rutkin, Aviva. *Crowdsourced Song lets the Masses Compose — One Note at a Time.* Available at: https://www.newscientist.com/article/dn28105-crowdsourced-song-lets-the-masses-compose-one-note-at-a-time/, August 2015.

Salvador, Stan, and Phillip Chan. "FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space." *Intelligent Data Analysis* 11, no. 5 (2007): 561–580.

Sargent, Gabriel, Stanislaw Raczynski, Frédéric Bimbot, Emmanuel Vincent, and Shigeki Sagayama. "A Music Structure Inference Algorithm Based on Symbolic Data Analysis." *Submission to MIREX competition at 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011.

Schafer, J. Ben, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative Filtering Recommender Systems." In *The Adaptive Web: Methods and Strategies of Web Personalization*, edited by Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, 291–324. Berlin: Springer-Verlag, 2007.

Schenker, Heinrich. *Der Freie Satz.* Vienna: UE, 1935.

Schepens, Wijnard, and Marc Leman. "Chronicle: XML-Representation of Symbolic Music and Other Complex Time Structures." In *Structuring Music through Markup Language: Designs and Architectures*, edited by Jacques Steyn, 99–118. Hershey: Information Science Reference, 2012.

Selfridge-Field, Eleanor. "Social Dimensions of Melodic Identity, Cognition and Association." *Musicae Scientiae* (London) Discussion Forum 4A (2007): 77–97.

Selway, Anna, Hendrik Vincent Koops, Anja Volk, David Bretherton, Nicholas Gibbins, and Richard Polfrem. "Explaining Harmonic Inter-Annotator Disagreement using Hugo Riemann's theory of 'Harmonic Function'." *Journal of New Music Research* 49, no. 2 (2020): 136–150.

Shardanad, U, and P Maes. "Social Information Filtering: Algorithms for Automating Word of Mouth." In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence* (Madison), 1995.

Shepard, Rodger. "Circularity in Judgements of Relative Pitch." *Journal of the Acoustical Society of America* 36, no. 1 (1964): 23–46.

Slaney, Malcolm, Santa Clara, Kilian Weinberger, and William White. "Learning a Metric for Music Similarity." In *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference* (Philadelphia), 2008, 313–318.

Smith, Charles J. "Musical Form and Fundamental Structure: An Investigation of Schenker's 'Formenlehre'." *Music Analysis* 15, nos. 2/3 (1996): 191–297.

Smith, Jordan B., J. Ashley Burgoyne, Ichiro Fujinaga, David De Roure, and J. Stephen Downie. "Design and Creation of a Large-Scale Database of Structural Annotations." In *Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference* (Miami), 2011, 555–560.

Smoliar, Stephen W. "A Computer Aid for Schenkerian Analysis." *Computer Music Journal* 4, no. 2 (1980): 41–59.

Smoliar, Stephen W., Robert E. Frankel, and Stanley J. Rosenschein. "A LISP-Based System for the Study of Schenkerian Analysis." *Computer and the Humanities* 10 (1976): 21–32.

———. "Schenker's Theory of Tonal Music — its Explication Through Computational Processes." *International Journal of Man-Machine Studies* 10, no. 2 (1978): 121–138.

Soafer, S. "Qualitative Methods: What are they and Why use them?" *Health Services Research* 35, no. 5 (199): 1101–1118.

Song, Yading, Simon Dixon, and Marcus Pearce. "A Survey of Music Recommendation Systems and Future Perspectives." In *9th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, 19–22. June. London, 2012.

Spitz, Bob. *The Beatles: the Biography.* London: Little, Brown / Company, 2005.

Stanyek, Jason. "Forum on Transcription." *Twentieth-Century Music* 11, no. 1 (2014): 101–161.

Stevenson, John. *Capturing Similarity in Music.* http://www.city.ac.uk/news/2015/june/capturing-similarity-in-music. (Accessed: 18.06.2016). City University London, 2015.

Stewart, Osamuyimen, Juan M. Huerta, and Melissa Sader. "Designing Crowdsourcing Community for the Enterprise." In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 2009, 50–53.

Steyn, Jacques. "Framework for a Music Markup Language." *Proceedings of the 1st International Conference of Musical Applications using XML (MAX)*, 2002, 22–29.

Tagg, Phillip. "Analysing Popular Music: Theory, Method and Practice." *Popular Music* 2, no. 1 (1982): 37–67.

Thom, Paul. *The Musician as Interpreter*. Pennsylvania: The Pennsylvania State University Press, 2007.

Tschinder, Charlotte. "Automating Music Similarity Analysis in 'Sound-Alike' Copyright Infringement Cases." *Entertainment, Arts and Sports Law Journal* 25, no. 2 (2014): 60–68.

Turnbull, Douglas, Luke Barrington, and Gert R. Lanckriet. "Five Approaches to Collecting Tags for Music." In *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference* (Philadelphia), 2008, 225–230.

Turner, Tom. "What is Metadata?" *Kaleidoscope* 10, no. 7 (2002): 1–3. doi:10.1007/978-1-4020-6869-0_2.

Tymoczko, Dmitri. "A Computer Aid for Schenkerian Analysis." In *Proceedings of the 2nd international conference of Mathematics and Computation in Music* (New Haven), 2009, 258–273.

———. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. New York: Oxford University Press, 2011.

Typke, Rainer, Marc den Hoed, Justin de Nooijer, Frans Wiering, and Remco C. Veltkamp. "A Ground Truth for Half a Million Musical Incipits." *Journal of Digital Information Management* 3, no. 1 (2005): 34–39.

Typke, Rainer, Frans Wiering, and Remco C. Veltkamp. "Transportation Distances and Human Perception of Melodic Similarity." *Musicae Scientiae* 11, no. 1 (2007): 153–181.

Volk, Anja, W. Bas de Haas, and Peter van Kranenburg. "Towards Modelling Variation in Music as Foundation for Similarity." In *Proceedings of the 12th International conference of Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (Thessaloniki), 2012, 1085–1094.

Volk, Anja, and Peter Van Kranenburg. "Melodic Similarity Among Folk Songs: An Annotation Study on Similarity-Based Categorization in Music." *Musicae Scientiae* (London) 16, no. 3 (2012): 317–339.

Wakefield, G. "Mathematical Representation of Joint Time-Chroma Distributions." In *Proceedings of the International Symposium on Optical Science, Engineering and Instrumentation* (Denver) 99 (1999): 18–23.

Warren, Carol A. B. "Qualitative Interviewing." In *Handbook of Interview Research*, edited by James A. Gubrium Jaber F. Holstein. California: SAGE Publications, 2001.

Wason, Robert W. "Music Practica: Music Theory as Pedgagogy." In *The Cambridge History of Western Music Theory*, 3rd, edited by Thomas Christensen, 46–77. Cambridge: Cambridge University Press, 2006.

Wasserman, Todd. *Coca-Cola To Help Maroon 5 Crowdsource a New Song.* Available at: https://mashable.com/2011/03/01/coca-cola-maroon-5, March 2011.

Welker, Robert. "Abstraction of Themes from Melodic Variations." *Journal of Experimental Psychology Human Perception and Performance*, 1982, 435–447.

Whalen, Michael. *The Death of the Music Genre (mostly. . .)* Available at: https://medium.com/@michaeljwhalen/the-death-of-the-music-genre-mostly-7039094302f, April 2018.

White, Christopher Wm, and Ian Quinn. "Chord Context and Harmonic Function in Tonal Music." *Music Theory Spectrum* 40, no. 2 (2018): 314–335.

Wieczoekowska, Alicja A. "Music Information Retrieval." Chap. 216 in *Encyclopedia of Data Warehousing and Mining*, 2nd, edited by John Wang, 1396–1402. Hershey: IGI Global, 2009.

Wiggins, Gerant. "Computer Representation of Music in the Research Environment." In *Modern Methods for Musicology: Prospects, Proposals, and Realities*, edited by Tim Crawford, 7–27. Farnham: Ashgate, 2009.

Winkler, Peter. "Randy Newman's Americana." *Popular Music* 7, no. 1 (1988): 1–26.

Winn, John C. *Way Beyond Compare: The Beatles' Recorded Legacy, Volume One, 1957-1965.* Crown Archetype, 2008.

Womack, Kenneth. *Maximum Volume: The Life of The Beatles Producer George Martin, The Early Years, 1926–1966.* Chicago Review Press, 2017.

Yapriady, Billy, and Alexandra Uitdenbogerd. "Combining Demographic Data with Collaborative Filtering for Automatic Music Recommendation." In *Proceedings of the 9th International Conference of Knowledge-Based Intelligent Information and Engineering Systems* (Australia), 2005, 201–207.

Yoshioka, Takuya, Testuro Kitahara, Kazunori Komatani, Tetsuya Ogata, and Hiroshi Okuno. "Automatic Chord Transcription with Concurrent Recognition of Chord Symbols and Boundaries." In *Proceedings of the 5th International Society for Music Information Retrieval (ISMIR) Conference* (Barcelona), 2004, 3373–3376.

Zhao, Yuxiang, and Qinghua Zhu. "Evaluation on Crowdsourcing Research: Current Status and Future Direction." *Information Systems Frontiers* 16, no. 3 (2014): 417–434.

Zibikowski, Lawrence. *Conceptualizing Music: Cognitive Structure, Theory and Analysis*. New York: Oxford University Press, 2006.