Routledge
Taylor & Francis Group

Journal of New Music Research

ISSN: 0929-8215 (Print) 1744-5027 (Online) Journal homepage: https://www.tandfonline.com/loi/nnmr20

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

Anna Selway, Hendrik Vincent Koops, Anja Volk, David Bretherton, Nicholas Gibbins & Richard Polfreman

To cite this article: Anna Selway, Hendrik Vincent Koops, Anja Volk, David Bretherton, Nicholas Gibbins & Richard Polfreman (2020) Explaining harmonic inter-annotator disagreement using Hugo Riemann's theory of 'harmonic function', Journal of New Music Research, 49:2, 136-150, DOI: 10.1080/09298215.2020.1716811

To link to this article: https://doi.org/10.1080/09298215.2020.1716811

Published online: 29 Jan 2020.

JOURNAL OF NEW MUSIC RESEARCH
2020, VOL. 49, NO. 2, 136–150
https://doi.org/10.1080/09298215.2020.1716811

Routledge
Taylor & Francis Group

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

Anna Selway[a]*, Hendrik Vincent Koops[b], Anja Volk[b], David Bretherton[a], Nicholas Gibbins[c] and Richard Polfreman[a]

[a]Department of Music, University of Southampton, Southampton, UK; [b]Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands; [c]Department of Electronics and Computer Science, University of Southampton, Southampton, UK

ABSTRACT

Harmonic transcriptions by ear rely heavily on subjective perceptions, which can lead to disagreement between annotators. The current computational metrics employed to measure annotator disagreement are useful for determining similarity on a pitch-class level, but are agnostic to the functional properties of chords. In contrast, music theories like Hugo Riemann's theory of 'harmonic function' acknowledge the similarity between chords currently unrecognised by computational metrics. This paper, utilises Riemann's theory to explain the harmonic annotator disagreements in the Chordify Annotator Subjectivity Dataset. This theory allows us to explain 82% of the dataset, compared to the 66% explained using pitch-class based methods alone. This new interdisiplinary application of Riemann's theory increases our understanding of harmonic disagreement and introduces a method for improving harmonic evaluation metrics that takes into account the function of a chord in relation to a tonal centre.

ARTICLE HISTORY
Received 1 March 2019
Accepted 20 December 2019

KEYWORDS
Harmony; Riemann; inter-annotator agreement; music information retrieval; automatic chord estimation

1. Introduction

Music transcription by ear relies heavily on subjective perceptions of musical structures (Jairazbhoy, 1977; Klapuri, 2006). The subjective nature of perception can lead to disagreements between annotators on what is the 'correct' transcription. Annotator disagreement (or inter-annotator disagreement) is exemplified in music-theoretical discourse on popular music, where 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. Using the ear to transcribe harmony creates many subjective attributes that are related to the specific relationships between the component pitches and their overtone partials (Klapuri, 2006). These disagreements are exemplified in the vast amounts of heterogeneous transcriptions available in online repositories, such as in the Ultimate Guitar repository, [1] and Chordify.[2]

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 Koops et al. (2019), the Rock Corpus introduced by De Clercq and Temperley (2011), and the dataset used by Ni, McVicar, Santos-Rodriguez, and De Bie (2013). The research involving these datasets commonly aims to find an empirical upper bound for inter-annotator agreement of harmonic annotations. However, research exploring harmonic disagreement between annotators is in its infancy. In music-theoretical studies, authors often focus on comparing their own analyses, such as the comparative analysis of the Rock Corpus by the two authors of De Clercq and Temperley (2011). Music Information Retrieval (MIR) studies often compare the analyses of a relatively small number of songs, for example, in the work of Ni et al. (2013).

The metrics used to measure annotator disagreement in MIR studies commonly focus on a pitch-class agreement, i.e. the amount of pitch-class overlap among the chord labels of the annotators for a particular segment. It can be argued that this agreement occurs at the lowest level of abstraction, meaning we are purely observing if the notes on the surface are the same. The origins

CONTACT Anna Selway alkm1g12@soton.ac.uk
*neé Anna Kent-Muller.

[1] https://www.ultimate-guitar.com/.
[2] https://chordify.net/.

of common-tone analysis are arguably music theories such as Hugo Riemann's theory of 'harmonic function' (Riemann, 1896, 1992). Therefore, by adopting Riemannian theory, we can utilise the harmonic function, which, enable us to establish links between chords intuitively perceived as similar in music (e.g. the relative major or minor) (Krumhansl, 1998; Krumhansl, Bharucha, & Kessler, 1982). This enables us to ascertain similarity that was not perceived in the current common tone approach.

This paper presents a new application of Hugo Riemann's theory of 'harmonic function', as a method of explaining some of the apparent disagreements in human-annotated datasets of harmonic transcriptions, specifically the Chordify Annotator Subjectivity Dataset (CASD). Through observing that Riemannian theory can explain some of the annotator disagreement present in CASD, it is shown that perceptual similarity is present between these chords. This agreement exists on a more abstract music-theoretical level, and can potentially show a higher level of agreement at this more musically informed level of harmonic function. This research provides a new application of Riemannian theory warranting exploration of the theory's relationship to music perception. MIR could benefit from adopting this music-theoretical approach to enable the evaluation of harmonic disagreement at a more abstract level, taking into account this perceptual similarity.

Synopsis. After introducing our music-theoretical approach in Section 1.1, we turn to a brief overview of the dataset used in this paper in Section 1.2 and a discussion of disagreement in Section 1.3. After introducing our methodology in Section 2, we present some example analyses in Section 3 and then provide the overall results on our application of Riemannian theory in Section 4. This paper closes with a discussion and conclusion in Section 5.

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 symbolic representation, such as sheet music, meaning transcribing harmony by ear is an often neglected skill. Aural transcription relies heavily on personal subjective perceptions – firstly in terms of assessing the auditory cues, and secondly in their translation into musical structures. This increases the propensity for subjective influence, leading to annotator disagreement.

One such example, of the disagreements that arise from aural transcriptions of harmony, can be seen in the literature surrounding the first chord of the Beatles song 'A Hard Day's Night'. Ever since it was recorded, music theorists, experts and amateurs have tried to unravel the sound into its respective pitches, contributing to its 'holy grail' status of 'one of popular music's great unsolved mysteries' (Pedler, 2003). The complex cluster chord, with no 'original' notated version, has been perceived as everything from a G major chord, with suspended 4th, added 7th, 9th or 11th (Hickey, 2010; Pedler, 2003; Spitz, 2005) along with inversions (Fujita, Hagino, Kubo, & Sato, 1993), to an F major chord with added G and D (Winn, 2008; Womack, 2017), to more complicated labels surrounding the function of the chord such as dominant 9th of F (Mellers, 1974), polytriad i7/5 in A♭ major (Porter, 1983), to a polychord which juxtaposes the tonic and subtonic (O'Grady, 1983). For further discussion of the variants in the analysis of this chord, we refer the reader to Koops (2019).

Music-theoretical approaches to harmony were developed for European tonal art music, although, researchers have debated in favour of its appropriateness as a method for analysing popular (and vernacular) music (Biamonte, 2010; Doll, 2007; Everett, 2004). For our music-theoretical approach, we have chosen to utilise Hugo Riemann's theory, due to the successful application of his theory in the work of popular music scholars, and our corpus being made up of songs using traditional triad-based harmony, thus featuring the tonic, subdominant and dominant chords or functions prominently (Biamonte, 2010, 2012; Capuzzo, 2004; Doll, 2007). Further research should consider utilising European art music (the genre the theory was designed for) to see if the genre the theory was written for also highlights a perceptual relationship between harmonic disagreement and Riemannian theory.

The 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 Agmon, 1995 on revisiting harmonic function using prototype theory). However, like many music theories, it was created with the concept of musical perception at its core (Clark, 2011; Riemann, 1896, 1992). Riemannian theory also lends itself well to our analysis as the method of 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 (Goldman, Jackson, & Sajda, 2018). This may, therefore, go some way to explaining the disagreements between annotators.

To enable us to use Riemannian theory, we required the musical score for each song in the dataset. 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. The scores are utilised in this paper for cues of important motives, alignment of the lyrics with the harmony and to provide the local and global keys. Though this is a limitation of this study, this paper aims to provide an impetus for future work that does not rely on the score, adapting Riemannian theory for the auditory domain. Using the scores has also enabled us to further explore some of the disagreements that cannot be explained by Riemannian theory. This includes disagreements over the level of granularity at which to annotate the harmony, along with observing musical features such as the simultaneous presence of two chords. This paper is not trying to place judgement over which harmony is correct, instead we wish to employ music-theoretical approaches to see how these could explain perceptual harmonic disagreements.

1.2. Dataset

To study disagreement in harmony transcriptions, this paper uses the Chordify Annotator Subjectivity Dataset (CASD) that was introduced by Koops et al. (2019). This dataset contains chord labels from four different professional annotators of 50 songs from the *Billboard* dataset. Burgoyne, Wild, and Fujinaga (2011) introduced the *Billboard* dataset, 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 (Bradlow & Fader, 2001). 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, according to their number of YouTube plays. At the time they were collected, the least-played song in the dataset had 76,000 plays and the most-played song over 13 million.

Koops et al. (2019) required their annotators' to have formal study of music and harmony at undergraduate or graduate level, experience in performing (for example in cover bands), and experience in transcribing popular music. This was to ensure high-quality transcriptions. 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 1 for an overview of the annotators. Half the annotators of this dataset were guitarists, and the other half were pianists; that is, all play chordal instruments. This is discussed further in Section 2.2.

To create CASD, the annotators' assignment 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 a 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, and was not set. Koops et al. (2019), however, found that all annotators asked for chords to be added, and the researchers maintained close contact with the annotators to enable them to request additions easily. The annotators were also given a chance to go back and edit their annotations throughout the study.

1.3. Disagreement

Koops et al. (2019) provide a detailed overview of the disagreement between annotators found in CASD. Altogether, it was found that each annotator used a particular set of chord labels – or vocabulary – for their transcriptions. Their vocabularies differed in size and content.

Table 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 number of chord labels per song.

Ann	Primary instrument	Years playing	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)

Difficulty is reported on a scale from 1 (easy) to 5 (hard). Standard Deviation is displayed in brackets.
Note: Table adapted from Koops (2019).

That is, in addition to sharing common 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 in which annotator was more likely to disagree (Koops et al., 2019).

Furthermore, in a pairwise analysis (i.e. a calculation of the average agreement between all possible pairs of annotators), it was found that annotators disagreed on 24% of the chord root notes. This increased with the complexity of chord labels to 41% when taking into account all pitch classes of the chords. The disagreement was even higher (an average of 46%) when inversions were taken into account. In a comparable experiment using annotations from formally trained musicians, Ni et al. (2013) reported around 10% disagreement among the annotators when compared to their consensus. Similarly, De Clercq and Temperley (2011) reported a 7.6% disagreement rate. In this paper, our disagreement rate is much higher, due to us not correcting any 'errors' made by the annotators. De Clercq and Temperley (2011) and Ni et al. (2013) claimed to correct errors that were unintentional before calculating their disagreement rate. We decided not to remove these as this paper did not aim to infer a right or wrong harmony but compare the disagreements between annotators, intentional or not.

Because we are interested in the sections with overall disagreement, we performed a global agreement analysis, instead of the pair-wise agreement performed by Koops et al. (2019). Global agreement refers to assessing if all annotators agree or disagree on the same chord. In our global agreement assessment, we observed when all four annotators agreed on the makeup of the chord. For this, we ignored embellishments such as sussed chords, 7ths, etc. This means, for example, we did not distinguish between C major, and C major7. This is because using Hugo Riemann's theory of 'harmonic function' (discussed in Section 2.1) these two chords would generally be perceived as having the same harmonic function because they share the same root 'C'. This worked for the music of our dataset, however, if the labelling of chords was for a genre of music such as Jazz, the use of embellishments is highly important, and our method of considering just the underlying chord could be too simplistic. Therefore, we performed our analysis at the major/minor level, meaning we did, for example, acknowledge a difference between C major and C minor. At this level, we found a global disagreement on 34% of the chords. In this paper, we will focus on this 34% disagreement, applying our music-theoretical approach discussed in Section 2, observing how much of this disagreement we can explain.

2. Method

To complete global analysis of the annotators' disagreement in CASD, the scores needed to be aligned with each annotator's audio-based annotations. This enabled us to not only observe the local keys for our functional analysis but also to observe whether there is a musical explanation for any remaining annotator disagreement. The scores were sourced from online repositories, such as *Musicnotes*.[3] This site was chosen due to its large catalogue (over 300,000 pieces), and its wide popularity as a source of sheet music. We successfully sourced sheet music for 41 out of the 50 songs in the dataset, thus our dataset consisted of only these 41 songs. The nine scores not available were mostly covers, mash-ups, or from a musical practice that features improvisation, and thus unlikely to be notated.

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 instead by professional arrangers (De Clercq & Temperley, 2011). Thus, it could be that these are only an approximation of the song, or to paraphrase the words of Cook (2005): a symbolisation of the musical sound rather than a representation, as it may not match the rhythm and notes exactly. These transcriptions themselves, therefore, may also suffer from the subjectivity of the transcriber. However, the scores do enable us 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 the CASD. To improve the annotations, we used human beat tracking data to correct the 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 produce subjectivity (Davies & Böck, 2014). To align the CASD annotations with the corrected beat annotations, we found for each score the closest matching beat in terms of physical time in the CASD chord label annotations. After repeating this process for each beat for each annotator, we obtained beat-corrected chord label annotations for each of the annotators in CASD. Once we had the information for

[3] *Musicnotes:* https://www.musicnotes.com/. The sites *Sheetmusicnow* (https://www.sheetmusicnow.com/) and *Sheetmusicplus* (https://www.sheetmusicplus.com/) were used when the scores were not available through *Musicnotes*.

which beat-per-bar our original chord labels fell on, we manually aligned our new per-beat annotations with the score's bar numbers.

For the remainder of Section 2, we will separate our methodology into two approaches. Firstly, we will discuss how we will utilise Hugo Riemann's theory of 'harmonic function' to explore annotator disagreement. Secondly, we will explore other ways of utilising the musical score to explore the remaining disagreement, such as prolongation and harmonic ambiguity.

2.1. Riemannian theory

One of the most prominent and influential theories of tonal harmony is introduced in the music-theoretical discourses of Riemann (1896, 1992). His theory moves away from the traditional harmonic emphasis on a triad's relationship to the tonic, instead of focusing on what Riemann describes as a functional theory of harmony. The functional theory of harmony is concerned with the harmonic purpose of a chord rather than with chord identity. Riemann's theory, he insisted, was founded on the now-discredited notion of harmonic dualism – which declares major and minor modes as being from nature (Bernstein, 2006). The system has, however, been shown to provide a useful concept of harmonic function. According to Riemann, there are three types of harmonic functions: the tonic (T), dominant (D) and subdominant (S) (Hyer, 2011). Importantly, these T, D and S functions are not only associated with the chords I, V, and IV, respectively. Riemann states how multiple chords can assert the same harmonic function through utilising a set of substitutions (Harrison, 1994). This means we can ascertain similarity through establishing which chords have the same harmonic function, i.e. which chords are 'substitutable'. Thus, these harmonic functions establish links between what would have been believed to be disparate, 'dissimilar', chordal structures, meaning two chords can have the same sense of function regardless of audible differences between them (Harrison, 1994). This would imply that, on a functional level, some assumed annotator disagreement has a perceptual similarity present between the chords. Or even, that the annotators made the annotations with a latent model of harmonic function in mind.

For this approach, we use the three basic harmonic substitutions; the *Variante*, *Parallele* and *Leittonsweschel* (Hyer, 2006, 2011), as a method to explain how some differences between annotators is perceptually an agreement. The first, the *Variante*, describes the major or minor substitutions of a chord with the same root but opposite mode. For example in Figure 1(a), C major (T) and C minor (t) are related by a *Variante* substitution – moving the third up or down a semitone (down to move major to minor and up for minor to major).

The second substitution defined by Riemann is the *Parallele*; this substitution is commonly known in English-language harmony literature as the 'relative' – connecting the major and minor triads whose roots are a minor third apart (manipulating a fifth of the chord up a tone from major to minor and in reverse the root moves down a tone). For example in Figure 1(b), C major (T) moves to A minor (Tp) by moving the fifth G up a tone to the root of A minor, A. In reverse, to substitute A minor (t) with C major (tP) the root of A minor (A) is moved down a tone to G to create C major. The final substitution discussed is the *Leittonsweschel*. This establishes a relationship between major and minor triads which have roots a major third apart (e.g. C major and E minor). In this substitution, the movement is applied to a different scale degree depending on whether the chord that is being substituted is major or minor. The root of a major chord moves down a semitone, whereas a minor chord's fifth moves up a semitone (Figure 1(c)). Through these substitutions Riemann argued that each different chord of a major or minor scale holds at least one possible function, though some chords, such as iii and vi, can hold multiple functions (Table 2).

Once we have associated substitution labels with each annotator chord label, we compare all annotators' substitutions for a beat. Those beats where there is some annotator disagreement (the 34% detailed in Section 1.3)

Table 2. The different chords that can hold each one of the three harmonic functions (tonic, dominant and subdominant), through different substitutions.

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)

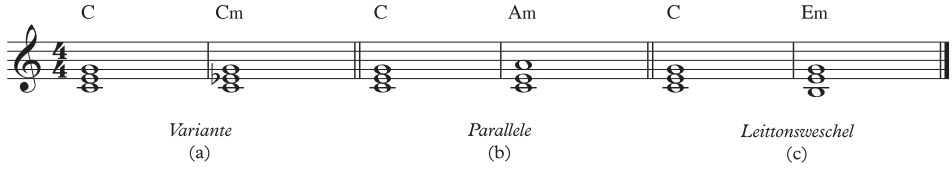


Figure 1. The three basic substitutions from Riemannian theory, *Variante*, *Parallele* and *Leittonsweschel*.

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 harmonic function. For example, assuming the key of C, the chord labels C, Am, Em, and C, would be analysed as `T`, `Tp`, `Tl`, `T`. Although they differ in their precise chord identity (`T`, `Tp`, `Tl`), they are all of the tonic function. The chord labels are different, but they are similar due to all being a substitution within the same function.

`Partial Agreement` refers to a majority agreement within the unique substitutions. We have utilised this category to enable us to account for subjectivity in harmonic annotation, allowing us to explore situations of majority agreement. For example, in C major, the chord labels Am, A, Cm, G are analysed as `Tp`, `TP`, `t`, `D`. 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 cannot be analysed under the same function and instead is analysed as having dominant function.

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

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 the category `No Agreement`. This is because the unique substitutions `T` and `D` are not of the same function. Though the reader can see a 75% agreement between the four annotators (three annotators perceiving the tonic and one the dominant), this approach was chosen as we were interested in whether we could explain the disagreements that currently cannot be explained in annotator disagreement. Current metrics already enable us to observe the similarity where there is a majority agreement on the same chord as the example discussed.

2.2. Score-based analysis

After completing the first stage of our methodology (identifying possible substitutions using Riemann's theory of 'harmonic function'), we utilised the already aligned scores to see if they could explain any of the remaining annotator disagreement. Firstly, we observed the disagreements in the harmony that were caused by different instruments playing different chords. This idea arose as the annotators were instructed to transcribe the complete harmony of the song in a way that, in their view, best matched their instrument. The annotators were split with two pianists and two guitarists (Table 1), thus we wished to explore whether the specific task, in combination with the annotator's primary instrument, could have caused some of the disagreement.

Secondly, we used the scores to observe if any remaining disagreement could be explained as disagreements on the level of granularity. For this, we utilised Heinrich Schenker's concept of prolongation. This term refers to the elaboration, or 'composing out', of music's underlying structure (named the Ursatz) (Cadwallader & Gagne, 2011; Drabkin, 2001). In music theory, prolongation is where a note governs a span of music without necessarily sounding (Drabkin, 2001; Forte & Gilbert, 1982; Pearsall, 1991). 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 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 (Forte & Gilbert, 1982; Larson, 1997).

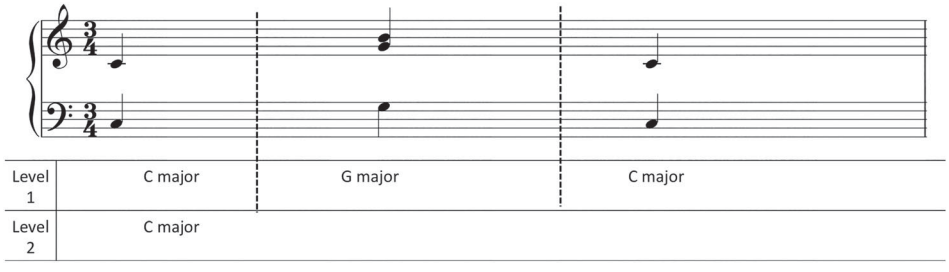


Figure 2. How a prolongation over a bar can affect harmonic annotation depending on which hierarchical level the annotator is observing their annotation at.

Larson (1997) discusses the audible perception of prolongations, and how different people may hear different levels of granularity, as found in Schenkerian analysis. The human ability to hear prolongations means that the displacement of traces also operates on various levels of the musical structure (Larson, 1997). Larson highlights how different listeners will hear different levels of prolongations; this may be based on musical training. In our approach, we did not perform full Schenkerian reductions to reduce the pieces down to their Ursätze. We did, however, observe within our disagreements the prominence of prolongations, and thus how different annotators may perceive the harmony at different levels of granularity, i.e. one annotator transcribing a harmony per bar, and another annotating per beat. For example, in Figure 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 less granular level, we can perceive the whole bar as being in C major.

3. Results: Example analyses

This section will provide five examples from a variety of songs in the dataset showing a mixture of the different categories of `Agreement`, `Partial Agreement`, and `No Agreement` (defined in Section 2.1). We will explore the harmonic disagreements of each extract not only using Riemannian theory, but also other methods of score-based analyses (as discussed in Section 2.2). Each example will first be discussed in terms of harmonic substitutions and the ability of that method to improve our understanding of disagreement, and then secondly we will discuss any remaining disagreements that can be explained through a feature of the score.[4]

3.1. 'All those years ago' by George Harrison

Our first example from CASD is an extract from 'All those Years Ago' by George Harrison. For this song, 77% of the disagreement between the annotators is explainable using Riemannian theory as the chords share the same function (`Agreement`), and a further 14% partially share the same function (`Partial Agreement`). Thus, in total we can explain 91% of the disagreement in this song, at least partially. Figure 3 shows an example of where we can fully explain the annotator disagreements using Riemannian Theory (`Agreement`). In the second half of bar 9 (the third and fourth sections of the diagram) 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, in fact, D major with the F♯ in the bass (`T`). Interesting A_2 perceives a single chord in bar 9, making us consider whether A_2 is transcribing at a different level of granularity to the other annotators. However, looking at bar 10 A_2 does not continue the pattern of annotating a single chord per bar, therefore, it is unlikely. However, they did not perceive the harmonic change the other annotators perceived in the second half of bar 9, just the introduction of a new root note. When looking at these chords without a function interpretation, the similarity between them is still apparent; all the annotators agree on the root (or bass note) being F♯ and the two chords share two common tones (out of four). Through using the Riemannian theory, we are able to explain how F♯minor7 is assuming a tonic function as the tonic *Leittonswechsel* (`Tl`) of D major. Thus, in this context, either chord would be capable of performing the same harmonic function.

Later in this example we have an agreement on the subdominant function, through more distantly related substitutions using diminished chords (second half of bar 10 in Figure 3). As explained in Section 2.1, diminished chords provide a dual function (often the 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 both an `S` substitution (due to the notes G and B♭ relating to G 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 the four annotators' chords have a subdominant function for their chosen harmony.

3.2. 'All through the night' by Cyndi Lauper

Cyndi Lauper's song 'All through the Night' shows another clear example of the category of `Agreement`. For this song, we can explain 68% of the disagreements using Riemann's functions (`Agreement`), and a further 6% through the partial matching of function (`Partial Agreement`). Figure 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 this 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 sees it as G major (`T`).

Finally, the same substitution is performed within the subdominant function in the second half of bar 50 (Figure 4), where A_1, A_2 and A_3 perceive it as in C major

[4] The examples were chosen for this section happen to show annotator 4 disagreeing frequently with the other annotators. As mentioned in Section 1.3, annotator 4 was no more likely to disagree with the other annotators.

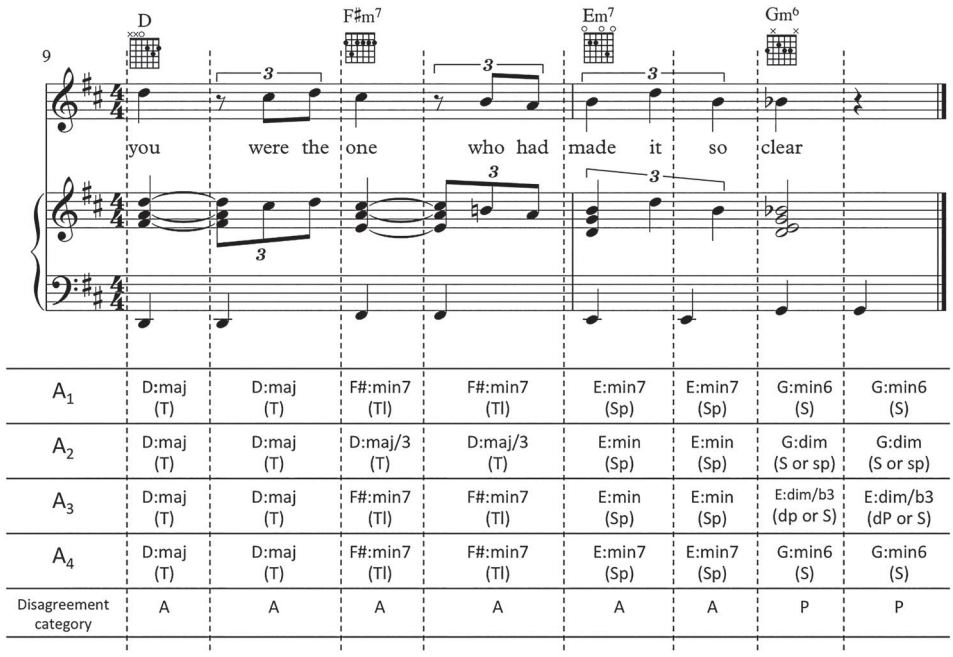


Figure 3. Bars 9 to 10 of 'All those Years Ago' by George Harrison. The figure shows the musical score aligned with each annotator per beat chord label. Here, we are interested in the use of the same harmonic function at the end of bar 9 and end of bar 10.

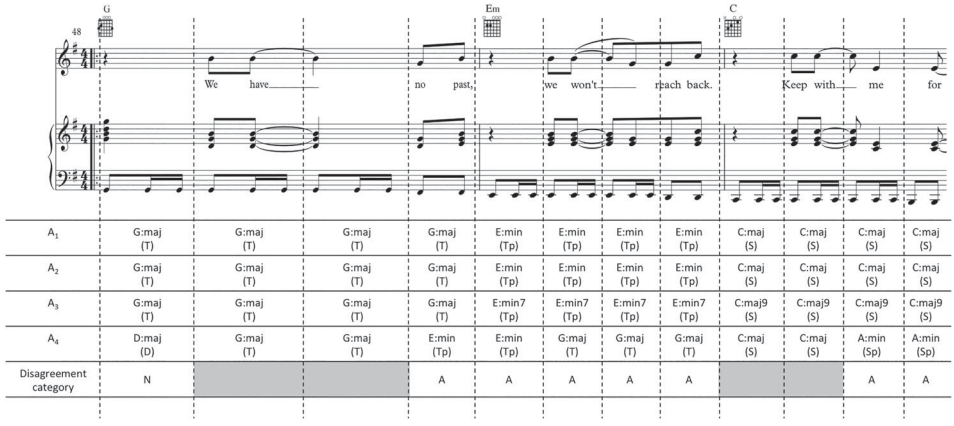
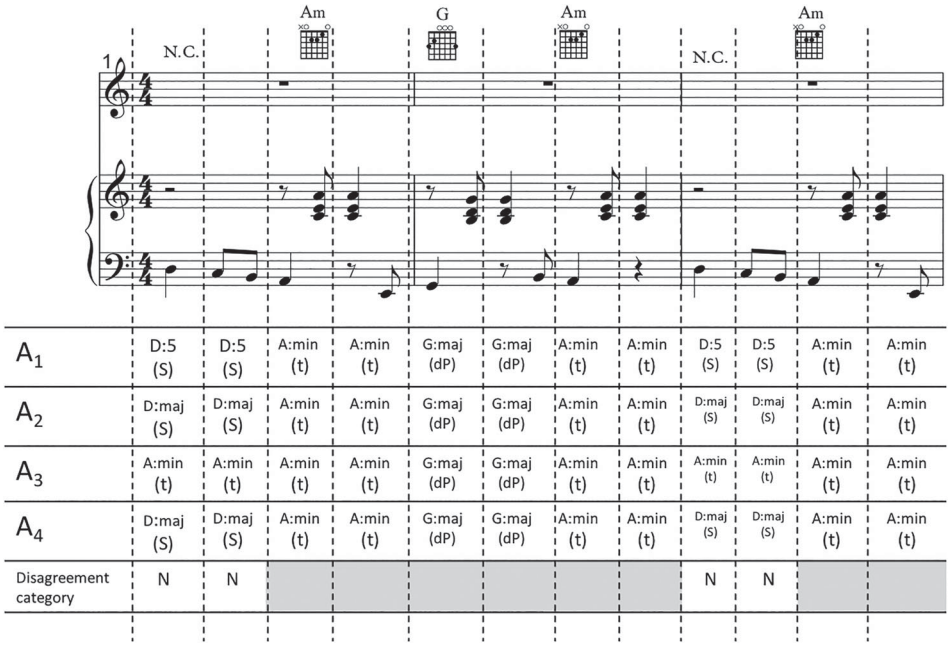


Figure 4. Bars 48–50 of 'All through the Night' by Cyndi Lauper. The figure shows the musical score aligned with each annotators per beat chord label. Here we are interested in the use of the same harmonic function at the end of bar 48, beats 2 to 4 of bar 49 and the second half of bar 50.

(S) and A_4 perceives the harmony as in A minor (Sp). This section, of 'All through the Night', also shows an example of where we cannot explain the annotator disagreement using Riemannian theory – bar 48. The first beat of bar 48 shows 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 could, however, be explained by the previous bar; bar 47 finishes with a D major chord, 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.

3.3. 'Super freak' by Rick James

Using Riemannian theory, we can only explain 8% of the harmonic disagreements in Rick James' song 'Super Freak' in the category of Agreement and a further 4% in the category of Partial Agreement. Riemannian theory cannot explain the harmonic disagreements (the No Agreement category) in this example (bars 1 and 3 of Figure 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)



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 5. Bars 1–3 of 'Super Freak' by Rick James. The figure shows the musical score aligned with each annotator's per beat chord label. Here, we are interested in the disagreement of the function of the chord at the beginning of bar 1 and 3.

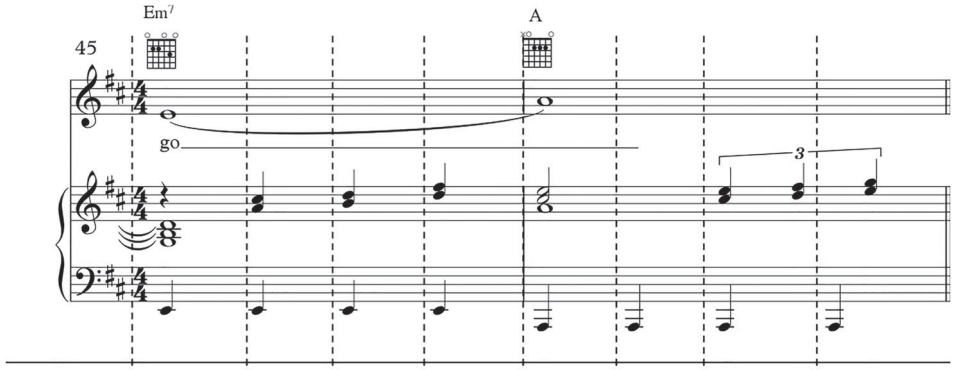
(with A_1 specifying the power chord D:5[5]) or A minor (t) as perceived by A_3. Thus, the annotators disagree on the function of the chord, between the subdominant and tonic functions.

The bass guitar riff at the beginning of bars 1 and 3 is repeated nearly continuously throughout the song, and this is a prominent feature of the piece. Through looking at the score, we can observe a few explanations for this disagreement. Firstly, the bass guitar part, as notated in the bottom stave (Figure 5), falls from the pitch D to an A, resembling a D major or minor chord (specifically a D:5). Then the piano part enters after the third beat with an A minor chord, which all the annotators agree on. 'N.C.' is notated above the stave over the guitar riff. This means 'no chord', suggesting no chord should be inferred because, strictly speaking, there is no harmony – it is a monophonic line. This means that the annotator who viewed the harmony as A minor, continuing the harmony of the proceeding and following beats, followed what this score implies. Using the score in this method enables us to explain a further 28% of the disagreements in this song in a similar way, explaining a total of 40% of the disagreements through both approaches.

[5] 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. This is a popular technique in popular music as the chord positioning can be easily transferred to multiple keys.

3.4. 'All those years ago' by George Harrison (revisited)

We return to the piece 'All those Years Ago' by George Harrison (the same piece used in Figure 3) for another example of `No Agreement` (see bars 45–46 in Figure 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 major/5 (T), whereas the other annotators perceive it still to be in E minor (Sp). As the functions are different, we cannot explain this using Riemannian theory. The same is apparent for bar 46, where A_4 perceives it to be D major/5 (T) and A_1, A_2 and A_3 perceive it as A major (D). However, in this example, we can observe the score in more detail to explain how some of these disagreements arose through different levels of granularity. For the fourth beat of bar 45, the annotators disagree between E minor and D major (with the differing functions of S and T). In the score, the middle stave has a rising third pattern, which raises up 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 reduced 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. A further 2% of the



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 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. Here we are interested in the disagreement in the function in bar 45, last beat, and bar 46.

disagreement in this song can be explained in this manner, through observing disagreements in granularity and the effect of Schenker's concept of prolongation.

4. Results: Statistics on Riemann

Overall, we find that the majority (39%) of the disagreement among the annotators' chord labels can be explained through substitutions of the tonic function. This aligns with the findings of De Clercq and Temperley (2011) and Burgoyne (2012) who both found that the tonic (in their case just `I`, not including its possible substitutions) was the most prominent chord in their corpus, followed closely by the subdominant, and finally the dominant. This also aligns with Biamonte's work on the 'stable tonic', 'less stable subdominant' and 'unstable dominant' as a way of generalising chord patterns in popular music (Biamonte, 2010, 2012). 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 De Clercq and Temperley (2011) who found the same dominance of the major mode in their *Rolling Stone* corpus. The next most frequent 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 Temperley (2011), Burgoyne (2012) and Biamonte (2010, 2012). Similarly to the tonic functions, most (28%) are of the major `S` function, while a much smaller number (6%) is explained by the minor `s`. Finally, 26% of the chord label disagreements can be explained through a dominant (`D` or `d`) function. We find again similar differences in occurrences of the two modes, with `D` amounting to 17%, and `d` explaining 9%.

The most commonly agreed upon chords in the `Agreement` subset,[6] that can be explained by Riemannian theory, are substitutions of the tonic function (like the whole dataset). Figure 7 shows the frequencies of tonic substitution co-occurrences in the `Agreement` subset. It highlights which substitutions are likely to be perceived when the majority of annotators agree on the substitution label on the Y -axis, what the other annotators are most likely to have perceived is displayed on the X -axis. This reveals which substitutions most often appear together in explaining the chord label disagreement between the annotators in the subset `Agreement` category. Within the tonic function, we find that `T` with `Tp`, `Tp` with `T`, and `T` with `Tl` are the most frequent

[6] When discussing the results of this dataset, the focus of the discussion will be on aspects of each category that can be explained using Riemannian theory. Therefore, in the categories `Agreement` and `Partial Agreement` we will focus on the same function chord labels. We will discuss disagreement that cannot be explained, and thus across function disagreement in reference to the category `No Agreement`.

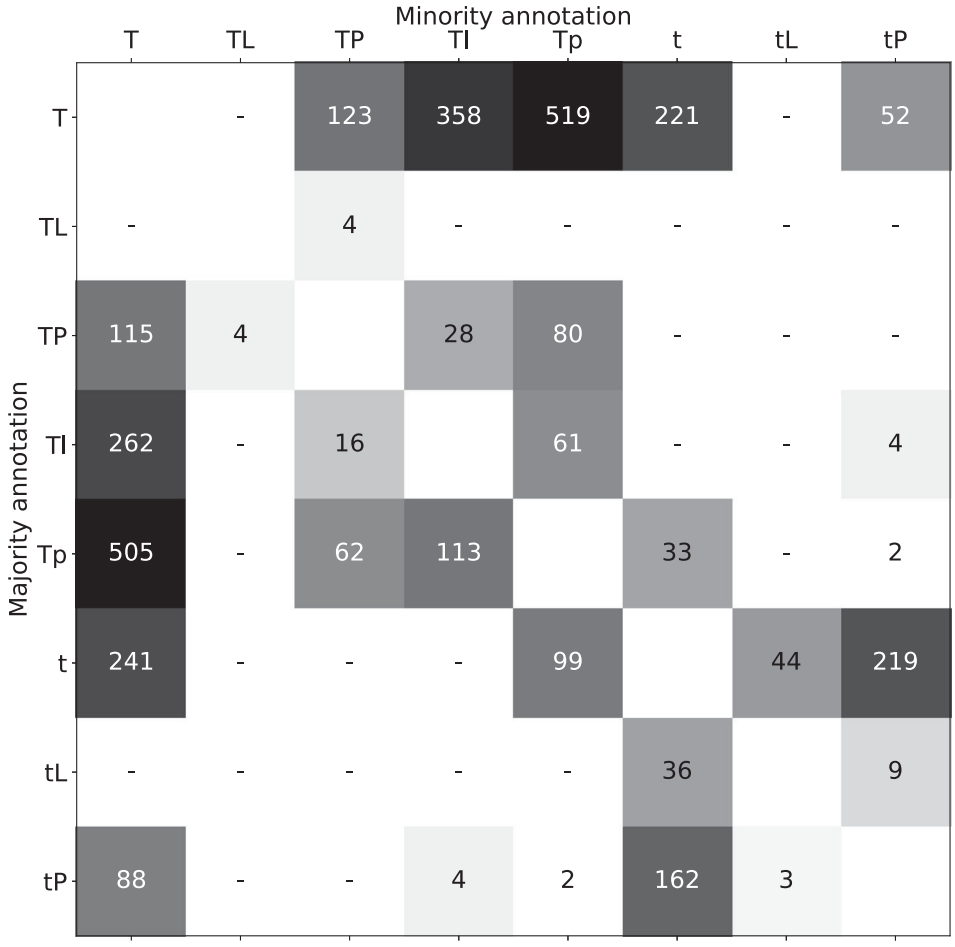


Figure 7. Frequencies of *tonic* substitution co-occurrences in the `Agreement` subset of the CASD. The numbers represent the frequency of co-occurrence of a majority substitution class (the most frequent substitution shared between the four annotators for a single beat) and other substitutions. The disagreements most often occur between *Parallele* and *Variante* related chords.

co-occurring substitutions. This shows that there is a strong perceptual confusion between `T` and `Tp`.

Unsurprisingly, we find the most likely chords to be confused are simple substitutions with their root function (e.g. `T` with `Tp` – chords that have two common tones, and are therefore very similar in their components). A small amount of literature has focused on the similarities between a chord and its *Parallele*; one such work comes from Krumhansl et al. (1982) who found a pattern of correlations reflecting a strong relationship between a major scale and its relative minor. The sheer prominence of using 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. This is again shown through the chord's two common tones, for example between C major (C,E,G) and A minor (A,C,E). However, the idea that we perceive this similarity, or that this type of similarity could cause us to confuse a chord that is 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 basic substitutions (*Variante*, *Parallele* and *Leittonswechsel*), the mode of the substitution chord changes. In this paper, we have found a vast number of occurrences where participants have disagreed on the mode. Thus, from an auditory point of view, there is a perceptual similarity between them. (Krumhansl, 1998) discusses the psychological reality of neo-Riemannian transformations (distinct from substitutions in terms of Riemannian theory of function).[7] Krumhansl (1998) highlights that the similarity between chords related by the *Leittonswechsel* transformation can be explained by the importance of pitch proximity and the fact that this requires the shift of a single note by just

[7] Unlike Riemannian theory, which replaces a chord with another, neo-Riemannian theory observes how chords progress to each other, horizontally, through a chord progression.

one chromatic step –this is the same for our *Leittonswechsel* substitution, where the chord requires a single pitch shift, thus retaining two common notes.

The most common subdominant functions to be disagreed upon can be seen in Figure 8. The most commonly disagreed upon chords are `S` with `Sp` and `Sp` with `S` – again showing the *Parallele* substitution to be most involved in explaining disagreement. Substantially less common is `S` with `Sl` (yet, still with greater frequency than other substitutions).

The most common dominant substitutions to be disagreed upon can be seen in Figure 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 hears a chord as having a dominant function rather than the pitches or specific chord. Therefore, that the listener hears chords as a function, rather than as a specific collection of pitches. We also find, as for the tonic and subdominant, that it is common for the *Parallele* substitution to be useful for explaining disagreement. However, the dominant also sees a prominence of the minor dominant being confused with its major *Parallele*, which we have not found for either of the other functions.

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

In contrast, within the `No Agreement` category, where we are looking at 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`. This shows that the most prominently occurring disagreements are between the basic functions, without any substitutions. The most common disagreements include the tonic function within the disagreement.

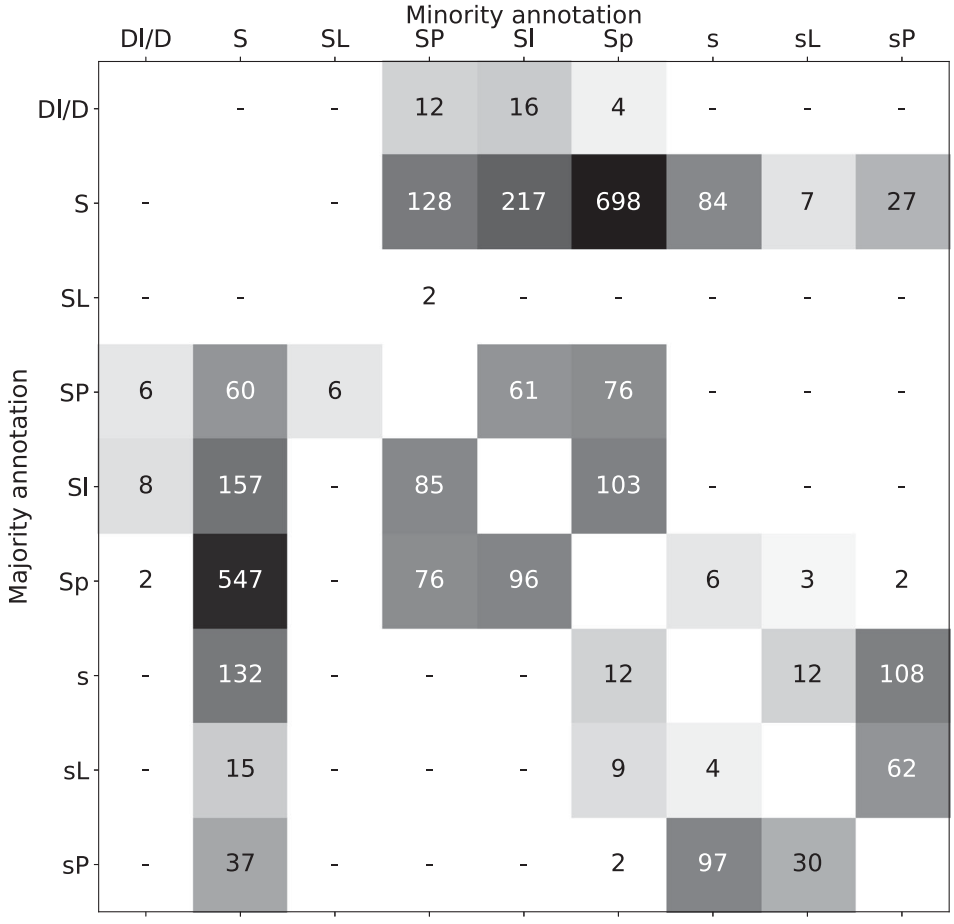


Figure 8. Frequencies of *subdominant* substitution co-occurrences in the `Agreement` subset of the CASD. Here, the disagreements most often occur between *Parallele* related chords.

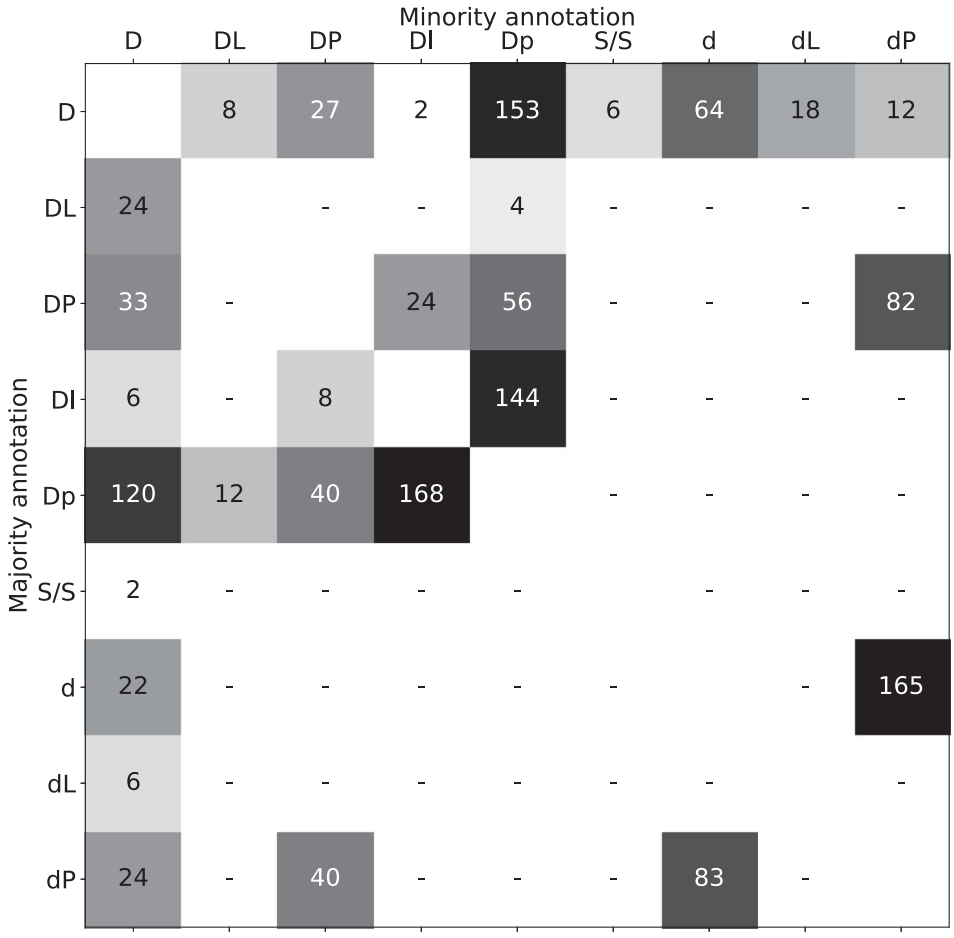


Figure 9. Frequencies of *dominant* substitution co-occurrences in the `Agreement` subset of the CASD. Here, the disagreements most often occur between the *Parallele* related chords, and substantially less often (but still worth noting) between *Variante* related chords.

Using our music-theoretical methods, we can explain, at the functional level, a total of 48% of the harmonic disagreements in CASD. Together with the sections in which there is full agreement (66%) between the annotators, this means that a little over 82% of the dataset can be explained. Firstly, using Riemann's theory of 'harmonic function', we can explain in full (the `Agreement` category) 27% of the disagreements in this dataset, and a further 13% partially (the `Partial Agreement` category), totalling 40% that can be explained using this method. We can also explain a further 5% through observing a disagreement over a chord caused by prolongation and granularity disagreements. Finally, we can explain another 3% of the annotator disagreements through the score, by highlighting ambiguities (for example the parts are having different harmonies).

5. Discussions and conclusions

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

Exploring the results, we found that the majority (40%) of disagreements among the annotators' chord labels can be explained through substitutions of the tonic, followed by the subdominant (34.8%) and then the dominant (25%). As discussed, these results align with previous work on popular music corpora (Burgoyne, 2012; De Clercq & Temperley, 2011) and music-theoretical explorations of popular music harmony (Biamonte, 2010, 2012). Observing in detail our `Agreement` category, we found that annotator

disagreement was most frequently explained through a disagreement between a root 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. It was highlighted how this is likely to be because chords related by one substitution will have two common notes, and are related through a single pitch shift (Krumhansl, 1998). However, the idea that therefore we perceptually hear a similarity between these chords related by one substitution is currently speculative. The results of this paper provide an impetus, warranting the exploration of Riemannian 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. It can be argued that these methods paint too bleak a picture of 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, if analysed in the key of `C`, Riemannian theory reveals that these differing chords can be easily explained, as both chords fulfil a tonic function. An initial analysis of CASD shows that within the part of the dataset that can be fully explained using our music-theoretical approach (`Agreement`), we only find around 49% root note agreement, and even less (39%) agreement on the root and the third. This means that there is a large difference in the notion of chordal agreement between the two approaches. A comparison of pitch-class oriented methods with our 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 the inter-annotator disagreement can be applied to datasets which include contributions from non-musically trained individuals (such as crowd-sourced harmony datasets). It would be worth exploring if this musically trained way of thinking about harmony is relevant to the general population, or is only applicable to those who have this musical training.

It is worth noting the limitations of our study. Our analysis was completed on a dataset containing diverse popular music and annotators, but pales in comparison to the amount of transcriptions found in online repositories, which raises questions on the ability to generalise our 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 dataset to investigate these properties with statistical validity is time-consuming and costly. It is also important to note that our current methodology requires a musical score to analyse according to Riemann's theory of 'harmonic function' (to enable us to determine any key changes within the music etc.). Due to the recording practices of popular music, the score is often a (subjective) transcription itself.

Our results show that some assumed annotator disagreement is actually 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. We believe that our results should inform future similarity measures used in music similarity tasks. Furthermore, our results could provide impetus to extend current evaluation measures of computational harmony tasks such as ACE, for example, by re-framing the task as a multi-label classification task, in which the correct chord labels share a common function. With the growing number of studies into annotator disagreement, we believe that is inevitable that computational harmony analysis will move towards modelling the perceived (or subjective) harmony of multiple annotators.

Acknowledgements

We thank W. Bas de Haas and Jonathan Driedger for their feedback on an early draft of this paper. We also wish to thank the Women in Music Information Retrieval (WiMIR) programme for the collaboration opportunity that led to this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Engineering and Physical Sciences Research Council [EP/L016117].

References

Agmon, E. (1995). Functional harmony revisited: A prototype-theoretic approach. *Music Theory Spectrum*, *17*(2), 196–214.

Bernstein, D. W. (2006). Nineteenth century harmonic theory. In T. Christensen (Ed.), *The Cambridge history of western music theory* (3rd ed., pp. 778–817). Cambridge: Cambridge University Press.

Biamonte, N. (2010). Triadic modal and pentatonic patterns in rock music. *Music Theory Spectrum*, *32*(2), 95–110.

Biamonte, N. (2012). Modal function in rock and heavy metal music. In M. Ayari, J. M. Bardex, & X. Hascher (Eds.),

L'analyse musicale aujourd'hui (pp. 275–290). Strasbourg: Delatour France.
Bradlow, E. T., & Fader, P. S. (2001). A Bayesian lifetime model for the Hot 100 Billboard songs. *Journal of the American Statistical Association*, *96*(454), 368–381.
Burgoyne, J. A. (2012). *Stochastic processes and database-driven musicology*. Montreal: McGill University.
Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). *An expert ground truth set for audio chord recognition and music analysis*. Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR), Miami, FL (Vol. 11, pp. 633–638).
Cadwallader, A., & Gagne, D. (2011). *Analysis of tonal music: A Schenkerian approach*. New York, NY: Oxford University Press.
Capuzzo, G. (2004). Neo-Riemannian theory and the analysis of pop-rock music. *Music Theory Spectrum*, *26*(2), 177–200.
Clark, S. (2011). On the imagination of tone in Schubert's Liedesend (D473), Trost (D523), and Gretchens Bitte (D564). In E. Gollin & A. Rehding (Eds.), *The Oxford handbook of Neo-Riemannian music theories* (pp. 294–321). New York, NY: Oxford University Press.
Cook, N. (2005). *Towards the compleat musicologist?* The 6th International Society for Music Information Retrieval Conference (ISMIR), London.
Davies, M., & Böck, S. (2014). *Evaluating the evaluation measures for beat tracking*. Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan (pp. 637–642).
De Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music*, *30*(1), 47–70.
Doll, C. (2007). *Listening to rock harmony*. Columbia: Columbia University.
Drabkin, W. (2001). *Prolongation*. Oxford University Press. Retrieved from http://www.oxfordmusiconline.com/subscriber/article/grove/music/22408
Everett, W. (2004). Making sense of rock's tonal systems. *Music Theory Online*, *10*(4). Retrieved from https://mtosmt.org/issues/mto.04.10.4/mto.04.10.4.w_everett.html
Forte, A., & Gilbert, S. E. (1982). *Introduction to Schenkerian analysis*. New York, NY: W. W. Norton & Company.
Fujita, T., Hagino, Y., Kubo, H., & Sato, G. (1993). *The Beatles: Complete scores*. Milwaukee, WI: Hal Leonard Publishing.
Goldman, A., Jackson, T., & Sajda, P. (2018). Improvisation experience predicts how musicians categorize musical structures. *Psychology of Music*, *48*(1), 1–17.
Harrison, D. (1994). *Harmonic function in chromatic music: A renewed dualist theory and an account of its precedents*. Chicago, IL: The University of Chicago Press.
Hickey, A. (2010). *The Beatles in mono*. Retrieved from lulu.com
Hyer, B. (2006). Tonality. In T. Christensen (Ed.), *The Cambridge history of western music theory* (3rd ed., pp. 726–752). Cambridge: Cambridge University Press.
Hyer, B. (2011). What is a function? In E. Gollin & A. Rehding (Eds.), *The oxford handbook of Neo-Riemannian music theories* (pp. 92–139). New York, NY: Oxford University Press.
Jairazbhoy, N. A. (1977). The 'objective' and subjective view in music transcription. *Ethnomusicology*, *21*(2), 263–273.
Klapuri, A. (2006). Introduction to music transcription. In A. Klapuri & M. Davy (Eds.), *Signal processing methods for music transcription* (pp. 3–20). New York, NY: Springer.
Koops, H. V. (2019). *Computational modelling of variance in musical harmony* (Unpublished doctoral dissertation). Utrecht University.
Koops, H. V., de Haas, B., Burgoyne, J. A., Bransen, J., Kent-Muller, A., & Volk, A. (2019). Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research*, *48*(3), 232–252.
Krumhansl, C. L. (1998). Perceived triad distance: Evidence supporting the psychological reality of Neo-Riemannian transformations. *Journal of Music Theory*, *42*(2), 265–281.
Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. (1982). Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 24–36.
Larson, S. (1997). The problem of prolongation in 'tonal' music: Terminology, perception, and expressive meaning. *Journal of Music Theory*, *41*(1), 101–136.
Mellers, W. (1974). *Twilight of the gods; The music of the Beatles*. Michigan: Viking Press.
Ni, Y., McVicar, M., Santos-Rodriguez, R., & De Bie, T. (2013). Understanding effects of subjectivity in measuring chord estimation accuracy. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(12), 2607–2615.
O'Grady, T. J. (1983). *The Beatles, a musical evolution*. Boston, MA: Twayne.
Pearsall, E. R. (1991). Harmonic progressions and prolongation in post-tonal music. *Music Analysis*, *10*(3), 345–355.
Pedler, D. (2003). *The songwriting secrets of the Beatles*. London: Omnibus Press.
Porter, S. C. (1983). *Rhythm and harmony in the music of the Beatles*. New York: City University of New York.
Riemann, H. (1896). *Harmony simplified: Or, the theory of the tonal functions of chords*. London: Augener.
Riemann, H. (1992). Ideas for a study 'on the imagination of tone' (R. W. Wason & E. West Marvin, Trans.). *Journal of Music Theory*, *36*(1), 81117.
Spitz, B. (2005). *The Beatles: The biography*. New York, NY: Little, Brown and Company.
Winn, J. C. (2008). *Way beyond compare: The Beatles' recorded legacy, volume one, 1957–1965*. New York, NY: Crown Archetype.
Womack, K. (2017). *Maximum volume: The life of Beatles producer George Martin, the early years, 1926–1966*. Chicago, IL: Chicago Review Press.