



Perceptual and automated estimates of infringement in 40 music copyright cases

RESEARCH

YUCHEN YUAN**

CHARLES CRONIN

DANIEL MÜLLENSIEFEN 

SHINYA FUJII 

PATRICK E. SAVAGE** 

 ubiquity press

*Author affiliations can be found in the back matter of this article

**Equal contribution

ABSTRACT

Music copyright infringement lawsuits implicate millions of dollars in damages and costs of litigation. There are, however, few objective measures by which to evaluate these claims. Recent music information retrieval research has proposed objective algorithms to automatically detect musical similarity, which might reduce subjectivity in music copyright infringement decisions, but there remains minimal relevant perceptual data despite its crucial role in copyright law. We collected perceptual data from 51 participants for 40 adjudicated copyright cases from 1915–2018 in 7 legal jurisdictions (USA, UK, Australia, New Zealand, Japan, People’s Republic of China, and Taiwan). Each case was represented by three different versions: either full audio, melody only (MIDI), or lyrics only (text). Due to the historical emphasis in legal opinions on melody as the key criterion for deciding infringement, we originally predicted that listening to melody-only versions would result in perceptual judgments that more closely matched actual past legal decisions. However, as in our preliminary study of 17 court decisions (Yuan et al., 2020), our results did not match these predictions. Participants listening to full audio outperformed not only the melody-only condition, but also automated algorithms designed to calculate musical similarity (with maximal accuracy of 83% vs. 75%, respectively). Meanwhile, lyrics-only conditions performed at chance levels. Analysis of outlier cases suggests that music, lyrics, and contextual factors can interact in complex ways difficult to capture using quantitative metrics. We propose directions for further investigation including using larger and more diverse samples of cases, enhanced methods, and adapting our perceptual experiment method to avoid relying on ground truth data only from court decisions (which may be subject to errors and selection bias). Our results contribute data and methods to inform practical debates relevant to music copyright law throughout the world, such as the question of whether, and the extent to which, judges and jurors should be allowed to hear published sound recordings of the disputed works in determining musical similarity. Our results ultimately suggest that while automated algorithms are unlikely to replace human judgments, they may help to supplement them.

CORRESPONDING AUTHOR:

Patrick E. Savage

Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan; School of Psychology, University of Auckland, Auckland, New Zealand
patrick.savage@auckland.ac.nz

KEYWORDS:

copyright; similarity; perception; melody; audio; lyrics

TO CITE THIS ARTICLE:

Yuan, Y., Cronin, C., Müllensiefen, D., Fujii, S., & Savage, P. E. (2023). Perceptual and automated estimates of infringement in 40 music copyright cases. *Transactions of the International Society for Music Information Retrieval*, 6(1), 117–134. DOI: <https://doi.org/10.5334/tismir.151>

1. INTRODUCTION

Music copyright law protects the legal rights and interests of music creators and performers, but in some music copyright infringement cases, its application has caused bitter controversy. For example, the multi-million dollar judgment that Robin Thicke and Pharrell Williams' #1 hit of 2013 "Blurred Lines" infringed on Marvin Gaye's "Got To Give It Up" sent shock waves throughout the music industry that continue to affect copyright law today (cf. section 7.5 and Fishman, 2018 for further details). As litigation becomes more frequent, unjustified music copyright lawsuits not only inhibit music creativity but also waste millions of taxpayer dollars annually to cover the adjudication of these disputes, while the value of global music copyright has been estimated to be over US\$30 billion (Page, 2021).

The legal system and music industry could both benefit from automated methods that could reduce subjectivity in music copyright decisions, and several recent studies have proposed such automated methods (Müllensiefen & Pendsch, 2009; Robine et al., 2007; Savage et al., 2018; Selfridge-Field, 2018). While the accuracy of some algorithms has been tested against previous court decisions, our preliminary study (Yuan et al., 2020) tested against both algorithmic and perceptual data to determine how different musical and extra-musical factors interact in copyright law. However, this study used a small sample of 17 US and Japanese cases that had been specifically chosen to emphasize melodic features, whereas many copyright cases may involve substantial similarity of non-melodic features such as lyrics or timbre. Our goal for this paper was to expand and replicate our preliminary analysis using a larger sample that included copyright disputes involving both melodic and non-melodic features.¹

"Substantial similarity" and "protectable expression" are central concepts in copyright law, the understanding of which could potentially be supplemented through automated and/or perceptual analyses. Infringement determinations require not only that the defendant can be shown to have copied musical material, but that this copying of protected musical expression was so extensive that the two works (or at least parts of the two works) are substantially similar (*Lotus v. Borland*, 1995). Data on degrees of computed and/or perceived similarity can help to determine objective standards for how much copying is required to be considered "substantial". (For a simplified, practical overview of the process of pursuing a copyright claim under US or UK law, see Lock & O'Rorke, 2022).

Complicating infringement determinations is the issue of access. If a defendant can establish that he had no access to the plaintiff's work when creating the defending work, no degree of similarity between the two works will overcome the presumption that the defendant did not copy the complaining work. If, unlikely as it may

be, two authors independently create identical musical works, both works obtain the same copyright protection, regardless of the order in which they were created. Accordingly, in two of the cases used in this experiment, *Selle v. Gibb* and *Repp v. Webber*², despite the remarkable melodic similarities between the contested works, the defendants avoided liability by establishing how extravagantly unlikely it was that they could have had any knowledge of the complaining works.

Evaluating what is considered "protectable expression" is more qualitative and complex. Many musical aspects such as scales, certain rhythmic patterns, and timbres are considered to be such basic and commonplace musical ideas or techniques as not to be copyrightable. For example, many blues songs use the same blues scale, the same 12-bar harmonic progression, vocal styles and instrumentation, but copying these aspects is not considered copyright infringement. Instead, melody (i.e., the sequence of pitches) and lyrics have traditionally played dominant roles over other musical factors (Fishman, 2018; Selfridge-Field, 1998).

While melody has traditionally been dominant, the relative importance of melodic vs. non-melodic features has never been formally defined. Instead, US copyright law relies on the "lay listener test" (*Arnstein v. Porter*) – which relies on the overall impression of non-expert jury members rather than any specific musical features argued by experts – to determine whether the defendant appropriated something original belonging to the plaintiff. While there is no fixed definition for how this concept should be implemented, the following criteria used in Lund's (2011) perceptual experiment were adapted from real instructions given to juries:

To find music copyright infringement between plaintiff's and defendant's songs, you must find that the songs are substantially similar. Two works are substantially similar if the original expression of ideas in the plaintiff's (Song #1) copyrighted work and the expression of ideas in the defendant's work (Song #2) that are shared are substantially similar. Original expression are those unique aspects of [the] plaintiff's song that are not common or ordinary to the genre or to music generally. The amount of similarity must be both quantitatively and qualitatively significant, that is the defendant's song copied either a substantial portion of the original expression of the plaintiff's song, or copied a smaller but qualitatively important portion of the plaintiff's song. (Lund, 2011:158)

The key point to note is that the definition does not specify any specific musical features (e.g., melody, rhythm), but merely asks listeners to evaluate "those unique aspects of [the] plaintiff's song that are not common or ordinary to the genre or to music generally".

After asking lay listeners to rate infringement in pairs of melodies from two influential copyright cases (*Swirsky v. Carey* and *Gaste v. Kaiserman*), Lund (2011) showed that lay listeners can perceive and judge infringement in the same pair of melodies very differently when non-melodic aspects of the music (e.g., tempo, key, orchestration) are modified, even when the melodies themselves are unchanged. Such effects have led to much debate, since it means that juries might come to different conclusions depending on the form in which they are allowed to listen to music when applying the lay listener test. For example, a core issue in the “Blurred Lines” dispute (*Williams v. Gaye*) was whether the jury should be allowed to listen to a full audio recording including lyrics and background instrumentation of the complaining work, or whether it should only be exposed to realizations of the sheet music that was deposited with the US Copyright Office (Fishman, 2018).

To quantitatively compare the effects of melody, lyrics, and other factors, we designed a controlled experiment where we constructed versions of a disputed musical work containing the full audio (the original audio version including lyrics, melody, and other features such as instrumentation), melody only (pitches and rhythms in MIDI representation), and lyrics only (text representation). Because of the historical dominance of melody, we originally predicted that participants would most accurately match past decisions when presented with melody-only versions, and that automated algorithms based on melodic data would more accurately match past decisions than ones based on full-audio data. However, our preliminary study found that the participants and algorithms showed minimal differences between melodic and non-melodic conditions, with if anything slightly higher accuracy matching past legal decisions when listening to full-audio (Yuan et al., 2020).

The main goal of our study was to replicate and extend the results of our preliminary study of 17 cases (Yuan et al., 2020) with a larger and broader sample of all 40 cases matching our criteria available at the Music Copyright Infringement Resource (MCIR; Cronin, 2018),³ representing a more diverse range of disputed criteria and copyright jurisdictions. While this expanded sample is still small and unbalanced enough to be primarily exploratory, the inclusion of cases revolving around lyrics and other non-melodic factors allows us to attempt a more thorough exploration of the value of perceptual and automated estimates of music copyright infringement.

2. RELATED WORK

2.1 PERCEPTUAL EXPERIMENTS

In one previous experimental study, Lund (2011) used two past court cases (*Swirsky v. Carey* and *Gaste v.*

Kaiserman) and manipulated MIDI representations of the works to change aspects such as tempo, rhythm, and instrumentation. Lund found that such manipulations reduced the accuracy of participants’ judgments of copyright infringements even though it was assumed that such non-melodic features should not play a role in decisions. Lund argued that this demonstrated that the “lay listener test” was flawed because it relies on subjective listening to audio recordings that may differ in non-melodic aspects. However, this study only tested two cases, and did not compare full audio recordings with these MIDI representations, so it remains unknown whether listeners are in fact more accurate when listening to MIDI representations than when listening to full audio recordings.

2.2 AUTOMATIC ANALYSIS

Automatic analysis of musical similarity is a long-standing challenge in music information retrieval (Casey et al., 2008), and a number of studies have applied this to the domain of copyright infringement (Müllensiefen & Pendsch, 2009; Robine et al., 2007; Savage et al., 2018; Selfridge-Field, 2018; Yuan et al., 2020; see Cason & Müllensiefen, 2012 and Malandrino et al., 2022 for detailed review). Often, this application is to a small number of famous cases: for example, Robine et al. (2007) applied a string-matching algorithm to five famous cases (*Bright Tunes v. Harrisongs*; *Fantasy v. Fogerty*; *Heim v. Universal*; *Repp v. Webber*; *Selle v. Gibb*). At larger scale, Müllensiefen and Pendsch (2009) developed an algorithm for judging melodic similarity that compares the profile of successive pitch intervals in two disputed songs against each other, while weighting them against a database of comparable profiles from 14,063 pop songs using a formula for estimating perceptual salience. When they applied this algorithm to a database of 20 past music copyright decisions focused on melodic similarity from the Music Copyright Infringement Resource (MCIR) (Cronin, 2018), they found the best-performing version of their algorithm was able to accurately identify 90% (18/20) of past cases.

Savage et al. (2018) developed a Percent Melodic Identity (PMI) method for quantifying melodic evolution based on automatic sequence alignment algorithms used in molecular genetics to measure melodic similarity (Savage & Atkinson, 2015). When they applied this method to the same set of cases as Müllensiefen & Pendsch, it accurately predicted 80% (16/20) of cases, despite being a simpler method that didn’t require calibration to an existing database of popular songs.

Recently, Malandrino et al. (2022) applied a range of automated algorithms to a larger sample of 164 cases from the MCIR and tested these against 50 famous plagiarism cases involving melodic similarity. They found some algorithms were able to predict up to 88–90% of

decisions. They also found that their automated method helped to assist participants with perceptual judgments of a subset of six cases randomly chosen from their set of 50 cases.

While the related task of cover song detection has a long history of study in music information retrieval (Serrà et al., 2010; Yesiler et al., 2019), to our knowledge no audio similarity algorithms had been tested for their ability to evaluate copyright infringement until our preliminary study (Yuan et al., 2020). However, many general audio similarity algorithms have been evaluated through the Music Information Retrieval Evaluation Exchange (MIREX) competition. We thus chose the audio similarity algorithm implemented in Musly, an open-source library of audio music similarity algorithms, because it has consistently performed at or near the top of audio similarity algorithms as evaluated in MIREX (Flexer & Grill, 2016).

2.3 OUR PRELIMINARY STUDY

We conducted a preliminary study with a small dataset of 17 music copyright court cases (14 from the USA and 3 from Japan; Yuan et al. 2020). Two automated algorithms focused on melody (Percent Melodic Identity [PMI]) and rhythm/timbre (Musly), proved to be effective in determining music similarity, matching past legal decisions with identical accuracy of 12/17 cases (~70%). Experiments were also conducted to collect perceptual data from 20 human participants, but we found no significant differences between the three conditions where the disputed sections were edited to contain either full audio, melody only, or lyrics only, with participants matching past decisions in between 50–60% of cases in all three conditions.

Since our preliminary study had limited size and scope with a dataset of only 17 court decisions and perceptual ratings from only 20 participants, we wanted to expand the testing data by including more usable music copyright cases. In particular, we wanted to change the sampling method from focusing only on cases which have court decisions focused on melodic similarity, to also include disputes involving similarities in lyrics, rhythm, timbre, or other non-melodic features, in order to fairly compare the roles of melodic vs. non-melodic features in copyright infringement judgments. To increase diversity and cross-cultural generalizability, we also aimed to expand our sample to include more languages and countries.

3. DATASET OF MUSIC COPYRIGHT INFRINGEMENT CASES

We followed a similar sampling strategy to that originally employed by Müllensiefen & Pendzich (2009) and our preliminary study (Yuan et al., 2020) of choosing all music copyright infringement cases from the Music

Copyright Infringement Resource (MCIR) (Cronin, 2018) whose final court decisions and full audio recordings of both disputed musical works were available. However, when Müllensiefen and Pendzich (2009) chose their sample, the MCIR only included eligible cases up to 2004. Combined with their choice to only focus on disputes involving melodic similarity, their sample was restricted to 20 cases from 1976–2004. Our expanded sample now includes all 40 MCIR cases with final decisions available as of October 2021 involving substantial similarity (of either melodic or non-melodic features) where complete audio recordings of both works are available, with decisions ranging from 1915–2018 (Table 1).⁴ 17 of these 40 cases were analyzed in our preliminary study and thus their main copyright issue focused on substantial similarity of melodies. The similarity types of the newly introduced 23 cases were not limited to melodic similarity but also included similarities in lyrics, rhythm, timbre, etc.⁵ Many of the cases involved disputed similarities in more than one category of musical factors (e.g., both melodic and lyrical similarity). Note that in the study we only considered cases in which substantial similarity of original expression played a major role in the decision, regardless of the outcome, but did not involve cases focused on sampling or the copyrightability of the contested musical expression in the complaining work (e.g., cases where two songs were similar but these similarities were shared with public domain works). While sampling cases apply the same “substantial similarity of protected expression” rubric, they tend to involve complex issues of literal copying, and defenses of transformative fair use.

These 40 cases were collected from 7 legal jurisdictions with somewhat diverse languages used in the songs in order to increase cultural diversity in the dataset for further study on adaptability to music other than Western music: 30 cases from the USA (all English except for one sung in French), 3 from Japan (Japanese, except 1 sung in English), 2 from People’s Republic of China (Mandarin/simplified Chinese), 2 from United Kingdom (English), 1 from Australia (English), 1 from New Zealand (English), and 1 from Taiwan (Mandarin/traditional Chinese). Of the 40 cases, courts found no infringement in 22 cases, and infringement in 18.

3.1 MUSICAL STIMULUS PREPARATION

The disputed segments of the musical works (mean length: 25s; range: 3–54s) were presented in one of three different versions: full-audio (the recorded versions including all instrumental and/or vocal parts), melody-only (MIDI rendition of the pitches and rhythms of the main melody), and lyrics-only (lyrics shown as visual text, without any accompanying audio).⁶

If several consecutive sections were disputed, we chose the longest continuous disputed section. For instance, in *Francescatti v. Lady Gaga*, both the repeated words “Juda” and “Judas” in the choruses and melodies

NO.	YEAR	JURISDICTION	CASE	COMPLAINING WORK	DEFENDING WORK	INFRINGEMENT DECISION?
1	2018	US	Pharrell Williams, et al. v. Bridgeport Music, et al.	Got to Give it Up	Blurred Lines	Yes
2*	2018	JP	Harumaki Gohan v. Mori	Hachigatsu no Rainy [August Rainy]	M.A.K.E.	Yes
3	2017	NZ	Eight Mile Style v. New Zealand National Party	Lose Yourself	Eminem Esque	Yes
4	2017	US	Parker, et al. v. Winwood, et al.	Ain't That a Lot of Loving	Gimmee Some Lovin	No
5	2016	US	Bowen v. Paisley	Remind Me	Remind Me	No
6	2016	US	Joel McDonald v. Kanye West, et al.	Made in America	Made in America	No
7	2014	US	Rebecca Francescatti v. Stefani Germanotta [aka "Lady Gaga"]	Juda	Judas	No
8	2012	US	Vincent Peters v. Kanye West, et al.	Stronger	Stronger	No
9	2010	US	Currin, et al. v. Arista Records, Inc., et al.	I'm Frontin'	Frontin	No
10	2010	AU	Larrikin Music Publishing Ltd. v. EMI Songs Australia Pty Ltd.	Kookaburra Sits in the Old Gum Tree	<u>Down Under</u>	Yes
11	2009	US	Bridgeport Music, Inc. v. UMG Recordings, Inc.	Atomic Dog	D.O.G. In Me	Yes
12	2009	US	Samuel Steele v. Jon Bongiovi, et al.	Man I Really Love This Team	I Love This Town	No
13	2007	US	Lil' Joe Wein Music, Inc. v. Jackson	It's Your Birthday	In Da Club	No
14	2007	TWN	People v. Hu	Liangshan Love Song	Bye Bye My Lover!	Yes
15	2005	US	Vargas v. Pfizer	<u>Bust Dat Groove Without Ride</u>	Advertisement for Celebrex	No
16	2005	US	Positive Black Talk v. Cash Money Records	Back That Ass Up	Back That Azz Up	No
17	2005	PRC	Yong Wang v. Zhengben Zhu	Send My Comrade to Beijing	Farewell to the Red Army	No
18*	2004	US	Swirsky v. Carey	One of Those Love Songs	Thank God I Found You	Yes
19	2004	PRC	Apollo Inc. v. Coca Cola (China) Inc.	When the Sun Rises	Sunrise	Yes
20*	2003	US	Cottrill v. Spears	What You See is What You Get	What U See is What U Get, Can't Make You Love Me	No
21*	2003	JP	Kobayashi v. Hattori	Dokomademo Ikou [Let's Go Anywhere]	Kinenju [Memorial Tree]	Yes
22	2002	UK	Malmstedt v. EMI Records	Jenny and I	Sleeping in my Car	No
23*	2002	US	Jean et al. v. Bug Music	Hand Clapping Song	My Love is Your Love	No
24*	2000	US	Three Boys Music v. Michael Bolton	Love is a Wonderful Thing	Love is a Wonderful Thing	Yes
25*	1997	US	Repp v. Webber	Till You	Phantom Song	No
26	1996	US	Santrayll v. Burrell	Uh Oh	Pepsi Ad featuring "Hammer"	No
27*	1994	US	Fantasy v. Fogerty	Run Through the Jungle	The Old Man Down the Road	No

NO.	YEAR	JURISDICTION	CASE	COMPLAINING WORK	DEFENDING WORK	INFRINGEMENT DECISION?
28	1993	UK	EMI Music v. Papathanasiou	<u>City of Violets</u>	<u>Theme from "Chariots of Fire"</u>	No
29*	1991	US	Grand Upright v. Warner	Alone Again (Naturally)	Alone Again	Yes
30*	1990	US	Levine v. McDonald's Corp.	Life is a Rock (But the Radio Rolled Me)	McDonald's Menu Song	Yes
31*	1988	US	Gaste v. Morris Kaiserman	Pour Toi	Feelings	Yes
32*	1987	US	Baxter v. MCA, Inc.	Joy	<u>Theme from 'E.T.'</u>	No
33*	1984	US	Selle v. Gibb	Let It End	How Deep is Your Love	No
34*	1978	JP	Harry v. Suzuki	The Boulevard of Broken Dreams	One Rainy Night in Tokyo	No
35*	1978	US	Herald Square Music v. Living Music	Day by Day	<u>Theme for N.B.C.'s "Today Show"</u>	Yes
36	1976	US	MCA Music v. Earl Wilson	Boogie Woogie Bugle Boy	The Cunnilingus Champion of Co. C	Yes
37*	1976	US	Bright Tunes Music v. Harrisongs Music	He's So Fine	My Sweet Lord	Yes
38*	1976	US	Granite Music v. United Artists	Tiny Bubbles	<u>Hiding the Wine</u>	No
39	1964	US	Nom Music, Inc. v. Kaslin	A Thousand Miles Away	Daddy's Home	Yes
40	1915	US	Boosey v. Empire Music	I Hear You Calling Me	Tennessee, I Hear You Calling Me	Yes

Table 1 The 40 music copyright infringement cases analyzed, ordered by year of decision. All cases involve vocal songs except the seven underlined instrumental works. Detailed summaries, legal documents, and audio recordings for each case can be found at the MCIR by clicking on the relevant hyperlinks.

*Included in preliminary study (Yuan et al., 2020). Jurisdiction: AU = Australia; JP = Japan; NZ = New Zealand; PRC = People's Republic of China; US = United States of America; TW = Taiwan (Republic of China).

in the breakdown sections were argued to share similarities, and consequently we selected sections spanning from the chorus through breakdowns of both songs for stimulus preparation. If several different and non-consecutive sections were disputed, we combined them for the study. For example, in *Peters v. West*, the plaintiff complained about the lyrics and rhyme scheme used in the defendant's works, so we chose the sections of potential lyrical similarity and combined them into a single audio file for each song. If a disputed section repeated multiple times with undisputed sections in between (e.g., chorus was disputed but not verse), we selected only one disputed section and ignored the repeated ones. If longer sections were disputed in vague terms (e.g., groove, genre, style) but shorter sections were disputed with specific issues (e.g., melodic/lyrical similarities), we focused only on the specific sections.

For the melody-only condition, we transcribed the melodies from the original audio recordings. Usually, we selected the vocal soloist line as the melody. If the melody included vocal harmony, we chose the

top notes, but ignored background vocal harmony if there was a soloist. If there was no vocal melody and/or if an instrumental melody was disputed, we used whichever instrument was highest and/or most melodic. If the vocal melody of the verse and/or the chorus and an instrumental melody from another part were both disputed, we chose the lead vocal melody and predominant instrumental melody. In order to control for all non-melodic factors including instrumentation, key, and tempo, the transcribed melodies from the original audio recordings were edited as necessary to exactly correspond to the audio recordings: these transcribed melodies were transposed to have a tonic of C, and were then recorded using the MIDI piano in MuseScore, played back at a tempo that was the average of the tempi from the plaintiff and defendant recordings. A metronome sound was added to all transcribed melodies to support listeners to perceive the rhythms/meters of songs and avoid any confusion in the absence of the instrumental background that would normally provide the metric context.

For the lyrics-only condition, we found the lyrics from online resources or manually transcribed the lyrics when the lyrics were not available online. There were six court cases involving instrumental works without lyrics (cf. Table 1) – these were not included in the lyrics-only study.

4. PERCEPTUAL EXPERIMENT

4.1 EXPERIMENT DESIGN

We conducted an online perceptual experiment where participants were each asked to judge substantial similarity for the 40 cases.⁷ To avoid possible order effects caused by participants evaluating the same melody/lyrics multiple times, we split the experiment in two: one with only full-audio versions of musical works, and one with melody-only and lyrics-only versions presented to the participants. Each participant only completes one experiment (this means each participant is never exposed to the same melody or lyrics twice). In our preliminary study we also included randomly selected pairs of melodies whose relationships were not in dispute (i.e., comparing a defendant's work against a randomly selected plaintiff's work) as well as disputed pairs of melodies to ensure that perceptual judgments were more accurate than would be expected by random chance. However, because these demonstrated very clear non-random judgment (near 100% accuracy for random pairs, with much lower accuracy of ~50–60% for disputed pairs; Yuan et al., 2020) and because these random pairs doubled the length of the experiment, we removed the random pairing for the expanded experiment; that is, in the new experiments, the participants compared only the originally disputed pair of musical works presented as either full-audio, melody-only, or lyrics-only versions.

This gave a total of 40 different pairs of musical works to evaluate in the experiment with only full-audio, and a total of 74 (40 cases shown as melody-only + 34 cases shown as lyrics-only) different pairs of musical works to evaluate in the experiment with melody-only and lyrics-only versions. Note that the 6 cases involving instrumental works without lyrics were not included in the experiment with lyrics-only versions. The pairs of musical works were presented in fully random order in each experiment (without separate blocks for different conditions; e.g., any given sample might be melody-only or lyrics-only in the experiment with melody-only and lyrics-only versions). Each experiment took approximately 1–2 hours for one participant to complete evaluations for all pairs.

For each pair, the participant is given a pair of music excerpts, “A” and “B”. “A” is always a plaintiff's work while “B” is always a defendant's work. After listening to the full-audio or MIDI or reading the lyrics of the two music works and reading the jury instructions for the “lay listener test” used by Lund (2011; see introduction), the participant needs to answer two questions: 1) How

similar do you think A and B are? (5-point Likert scale: “not at all similar”, “a little similar”, “somewhat similar”, “very similar”, and “extremely similar”), and 2) Do you think B infringes on the copyright of A? (yes/no answer). When lyrics-only versions are shown, the participants are allowed to answer N/A for both questions if they are not able to make the judgment, specifying the reasons (e.g., cannot understand the foreign languages shown).

4.2 RESULTS

We collected perceptual data from 51 participants recruited by the first author (YY), 28 of whom did the experiment with only full-audio and 23 of whom did the version with melody-only and lyrics-only. All experiments were conducted online in English (during the COVID-19 pandemic) and all participants provided informed prior consent to participate in the study.

For the full-audio experiment, 14 participants were male and 14 were female. Ages ranged from 23 to 57, and the mean age was 36. The native languages of the participants were Chinese (26 participants), English (1), and Japanese (1). The 26 Chinese native speakers also reported language skills of English (26 participants), Cantonese (2), French (1), Japanese (1), Spanish (1), and Thai (1). The English native speaker reported Japanese language skills. The Japanese native speaker reported English language skills. 7 reported substantial music experience of more than 2 years while 21 had less than 2 years of musical experience.

In the experiment with melody-only and lyrics-only versions of musical works, 8 were male and 15 were female. Ages ranged from 22 to 58, and the average age was 39. The native languages of the participants were Chinese (22 participants), and Japanese (1). The 22 Chinese native speakers also reported language skills of English (22 participants), Cantonese (2), and Japanese (2). The Japanese native speaker reported English language skills. 4 reported substantial music experience of more than 2 years while 19 had less than 2 years of musical experience.

Figure 1 shows how accurately the 51 participants' judgment of infringement matched the official court decisions when they were given full-audio, melody-only, or lyrics-only versions of music pieces from the 40 court cases. Note that “accuracy” was defined as the proportion of participants agreeing with the court's decision, whether that decision was of infringement or no infringement.⁸ In Figure 1A individual data points represent mean accuracy for individual cases ($n = 40$) averaged across the 28 and 23 participants, while in Figure 1B, individual data points represent mean accuracy for individual participants ($n = 28$ for full-audio condition and $n = 23$ for melody-only and lyrics-only conditions) across the 40 cases.

The results in Figure 1 showing mean participant accuracy of 57.9%, 57.5%, and 51.5% for full-audio,

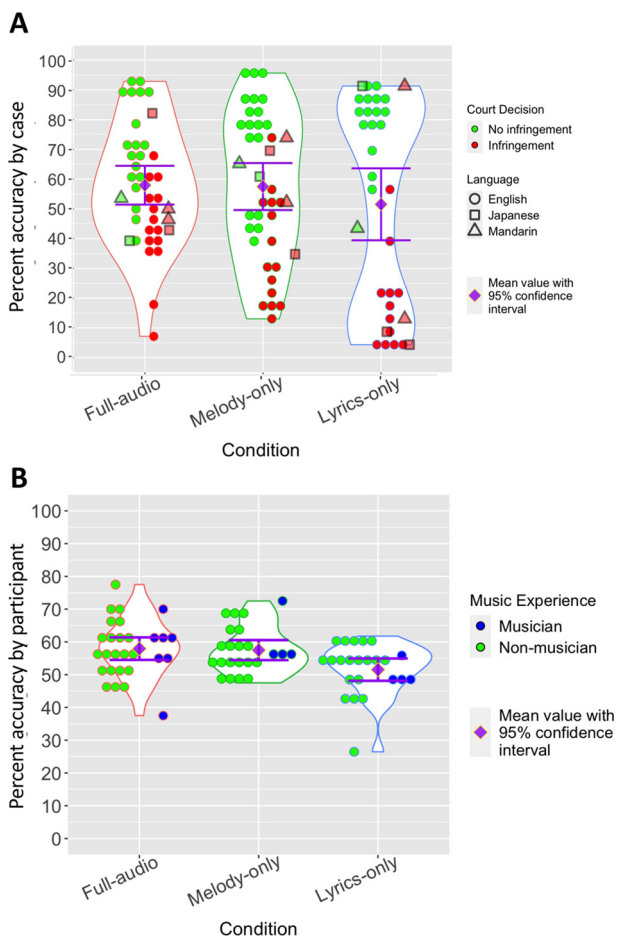


Figure 1 A) Accuracy of perceptual judgment for each of the 40 court cases, as measured by the percentage of the 51 (28 for full-audio, 23 for melody-only and lyrics-only) participants whose judgments of music copyright infringement matched court decisions. B) The same data plotted with the 51 participants as units averaged across court cases instead of vice-versa. Music experience of the participants is indicated by the filling color of dots; blue = self-reported musician; green = non-musician.

melody-only, and lyrics-only groups respectively are similar to the ones we found in our preliminary study with a smaller dataset and fewer participants (58%, 54%, and 49%, respectively). Variability is much greater in Figure 1A than Figure 1B, suggesting that the main source of variation comes from individual copyright cases, not the way different individuals perceive musical similarity. This point is also suggested by a comparison of musicians and non-musicians, who performed similarly (Figure 1B). No clear differences are notable for the small subsets of cases with Japanese or Chinese languages.

Notably, cases in all conditions show skewing toward higher accuracy for non-infringing cases (Figure 1A). This suggests that participants hold a high threshold for judging copyright infringement and tend to err on the side of no infringement. This raises the possibility that participants may not have been able to accurately predict infringement, with their performance instead due to chance and the slight baseline skew toward no infringement (22/40 cases).

To investigate this possibility we performed signal detection analysis using Receiver Operating Characteristic (ROC) curves for both mean perceived similarity and proportion of perceived infringement for all three conditions. The results confirmed that participants performed significantly above chance for full audio and melody-only conditions ($AUC > 0.7$, $p < .05$), but were at chance levels for the lyrics-only condition ($AUC < .5$, $p > .7$; Figure 2). These patterns were consistent whether based on perceived similarity or perceived infringement (Figure 2). Because lyrics-only data was thus effectively random, we exclude it from further analysis in the rest of the manuscript.

5. ALGORITHMIC ANALYSIS

We performed automatic similarity analysis of these cases using two different automated algorithms focused on melodic and audio similarity, respectively. Figure 3 shows the Receiver Operator Characteristic (ROC) curves for predicting these infringement decisions using different automated similarity thresholds.

5.1 MELODIC SIMILARITY (PERCENT MELODIC IDENTITY [PMI])

We chose the PMI (Percent Melodic Identity) method to calculate melodic similarity because it has been validated in previous research using a similar sample of copyright cases (Savage et al., 2018). Like Judge Learned Hand’s “comparative method” (Fishman, 2018) to test musical similarity, the PMI method begins by transposing two melodies transcribed in staff notation into a same key, eliminating rhythmic information by assigning all notes equal time values, and then aligning and counting the confluence of notes.⁹ Following the procedure, we prepared note sequences of disputed melodies all transposed to a C tonic for consistency (just as was done when preparing MIDI files). The PMI algorithm then automatically aligns each sequence pair using the Needleman-Wunsch global pairwise alignment algorithm (Needleman & Wunsch, 1970), and counts the number of identical notes (*ID*). The percentage of identical notes shared between melodies, named percent melodic identity (*PMI*) (Savage et al., 2018), is calculated by dividing *ID* by the average length of the melody pair (L_1 and L_2), as follows:

$$PMI = 100 \left(\frac{ID}{\frac{L_1 + L_2}{2}} \right)$$

5.1.1 Melodic Similarity Results

Receiver Operating Characteristic (ROC) analysis was used to assess the prediction given by PMI values. The area under the ROC curve (*AUC*) is 0.73 (Figure 3A). The optimal cutoff PMI value is 44.6% with sensitivity = 0.78 and specificity = 0.73. Using this cutoff, the PMI method

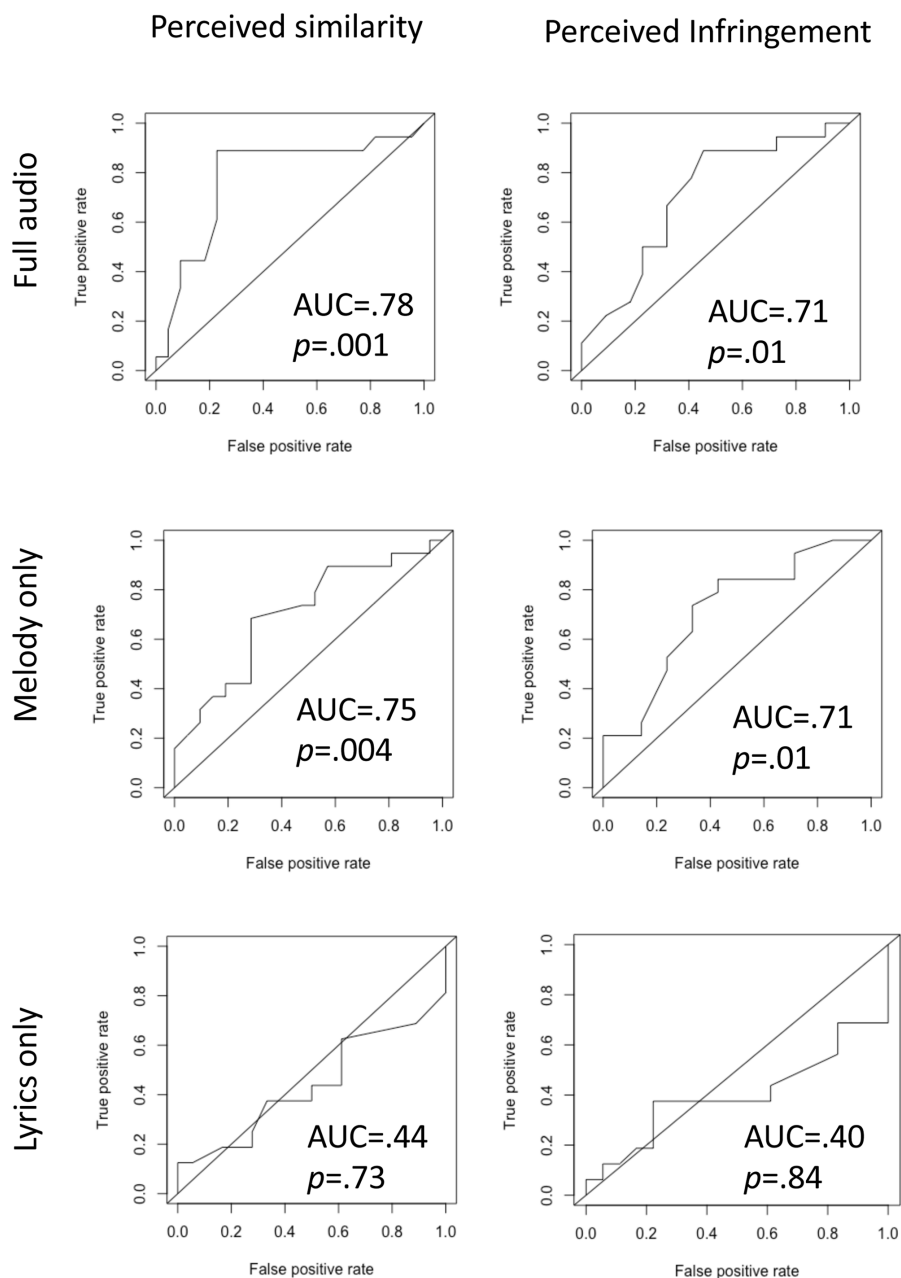


Figure 2 Receiver Operating Characteristic (ROC) curves for predicting copyright case decisions for the 40 cases based on mean perceived similarity (left column) and proportion perceived infringement (right column) for the three experimental conditions. AUC = Area Under Curve.

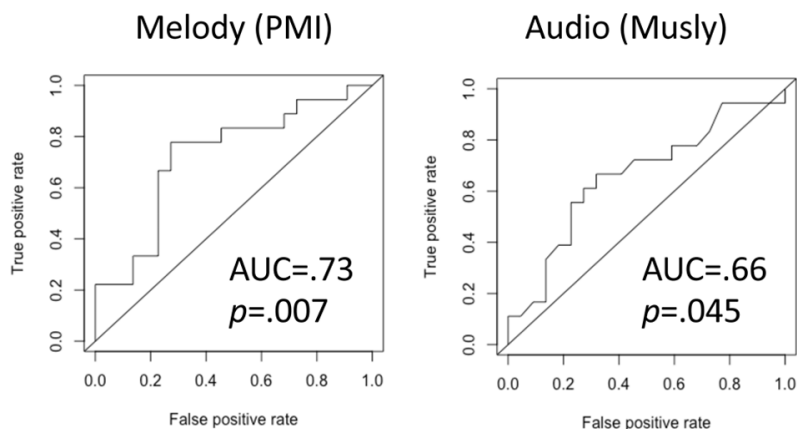


Figure 3 The Receiver Operating Characteristic (ROC) curves for predicting copyright case decisions for the 40 cases based on A) Percent Melodic Identity (PMI); and B) Musly’s audio similarity algorithm.

was able to accurately classify 30 out of the 40 cases (75%) to match their court decisions.¹⁰

5.2 AUDIO SIMILARITY (MUSLY)

Musly currently implements two music similarity algorithms. One implements the Mandel-Ellis audio similarity algorithm (Mandel & Ellis, 2005). The other one, which is the default one, improves the Mandel-Ellis algorithm to compute audio similarity for best results. Specifically, it computes a representation of each song's audio signal based on 25 Mel-Frequency Cepstral Coefficients (MFCCs) to estimate a Gaussian model and eventually a single timbre model to be compared, computes similarity between each pair of timbre models using Jensen-Shannon approximation, and normalizes the similarities with Mutual Proximity (Schnitzer, 2014; Schnitzer et al., 2011). We used the default algorithm because it has been found to have higher accuracy (Schnitzer, 2014).

We prepared the full-audio version of the music excerpts from the dataset of court cases and fed them to the default algorithm of Musly to compute similarity. The output of the Musly algorithm is a distance matrix where distances, i.e. differences, between every two songs are listed. Because the Musly default algorithm normalizes the results, all the distances range between 0 and 1. Consequently, we calculated the audio music similarity by subtracting distance values from 1 and multiplying by 100 to convert the results into percentage terms for consistency with our other methods.

5.2.1 Audio Similarity Results

The area under the ROC curve (AUC) is 0.66 (Figure 3B). The optimal cutoff threshold of Musly-calculated similarity is 43.0% with sensitivity = 0.67 and specificity = 0.68.

Using this cutoff, the Musly algorithm was also able to accurately classify 27 out of the 40 cases (68%) to match the court's decisions.¹¹

6. AUTOMATED VS. PERCEPTUAL JUDGMENTS

6.1 PMI VS. PERCEPTUAL DATA

Mean perceptual similarity of each court case was calculated by averaging participants' individual ratings of similarity. Figure 4A shows the relationship between PMI values and perceptual similarity under the three different conditions. Correlation analyses show that the PMI melodic similarity is significantly correlated with perceptual similarity for both full-audio and melody-only conditions (full: $r = 0.41$, $p = 0.009$; melody: $r = 0.77$, $p = 6.6 \times 10^{-9}$), but not for the lyrics-only condition ($r = 0.20$, $p = 0.26$).

6.2 MUSLY VS. PERCEPTUAL DATA

We also compared the Musly-calculated audio music similarity with the perceptual data collected. Figure 4B shows the correlation between Musly similarity and perceptual similarity of the 40 tested court cases under three different conditions for perceptual judgment. The Musly audio similarity measure has only weak, non-significant correlations with perceptual similarity for all three condition groups of "full-audio", "melody-only", and "lyrics-only" (full: $r = 0.065$, $p = 0.69$; melody: $r = 0.039$, $p = 0.81$; lyrics: $r = -0.21$, $p = 0.23$). Musly thus shows both weaker predictive power when compared to the actual court decisions and weaker correlations with participants' perceptual ratings than PMI.

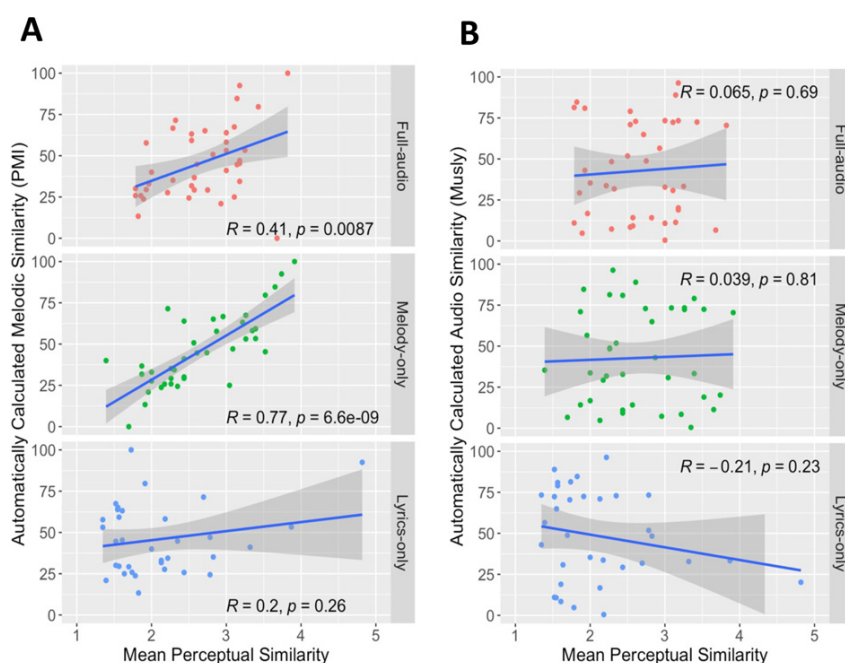


Figure 4 Mean perceptual similarity vs. automatically calculated for full-audio, melody-only, and lyrics-only conditions for the 40 cases for: **A)** melodic similarity (PMI); and **B)** audio similarity (Musly).

6.3 MULTIPLE LOGISTIC REGRESSION ANALYSIS INCLUDING AUTOMATED AND PERCEPTUAL VARIABLES

To determine which feature(s) best predicted copyright judgments, we ran a multiple logistic regression using variables that showed significant predictive power for court decisions in the signal detection analyses in Figures 2–3 (i.e., excluding lyrics-only data where participant judgments were at chance level and no algorithm for textual similarity was employed). Because mean perceived similarity and portion perceived infringement were highly correlated ($r = 0.93$), which creates problems with multi-collinearity, we only used the similarity ratings in the regression model, as they are more closely related to court decisions than the infringement judgements (cf. Figure 2). We thus included four variables in our regression:

- 1) “Perceived audio similarity” (mean averaged across the 28 audio-only participants for each of the 40 cases)
- 2) “Perceived melodic similarity” (mean averaged across the 23 melody-only participants for each of the 40 cases)
- 3) “Algorithmic audio similarity” (as calculated for each of the 40 cases using the Musly algorithm, as described in Section 5.2)
- 4) “Algorithmic melodic similarity” (as calculated for each of the 40 cases using the percent melodic identity [PMI] algorithm, as described in Section 5.1)

For completeness, we ran both frequentist and Bayesian models once using the full sample of $n = 40$ cases and once replicating the sub-sample of $n = 17$ cases used in the Yuan et al. (2020) preliminary study, with and without variable selection. The main results are summarized in Table 2 (cf. Supplementary Materials Tables S1–5 for alternative analyses). While the sub-sample of $n = 17$ cases was too small to draw strong conclusions from, all analyses provided qualitatively consistent results confirming the results of the ROC analyses suggesting that human perceptual ratings of similarity listening to full-audio excerpts best predicted the outcome of court cases. This variable demonstrated the highest standardized regression coefficients after controlling for

the other features through multiple logistic regression ($z = 2.0, p = .04$; Table 2). In contrast, perceived melodic similarity consistently predicted decisions less well than perceived audio similarity, and was excluded by the stepwise AIC (Akaike Information Criterion) variable selection procedure (Table 2).

7. QUALITATIVE ANALYSIS OF EXAMPLE CASES

To explore the qualitative dynamics underlying our results, we visualize our 40 cases based on their best-performing perceptual and automated features and examine contextual factors underlying the more extreme or surprising results. For this, we plot in Figure 5 the best-performing human perceptual ratings (perceived similarity listening to full audio) against the best-performing algorithmic predictor (PMI ratings of melodic similarity). This allows us to visualize and analyze extreme

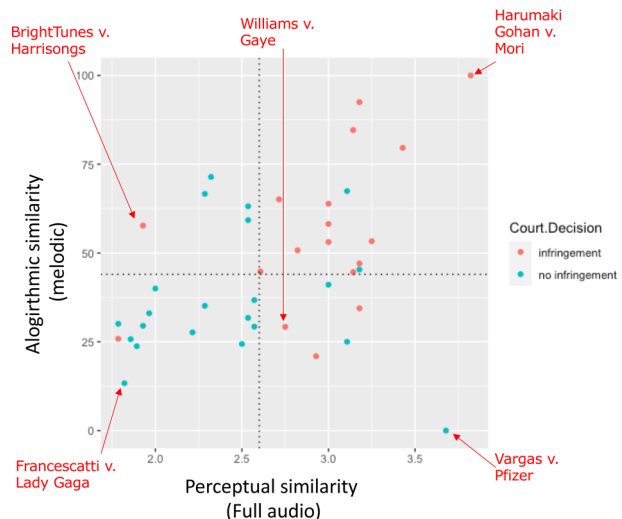


Figure 5 Scatterplot of the best-performing predictors of past copyright infringement for perceptual data (perceptual similarity with full audio: x-axis; 1–5 scale) and automated algorithms (Percent Melodic Identity [PMI] measure of melodic similarity: y-axis; 0–100% scale). Cases judged to infringe copyright are plotted in red, with non-infringing cases plotted in blue. Optimal cutoffs from the Receiver Operating Characteristic (ROC) curves (cf. Figures 2–3) are shown using dashed lines. Example cases representing extreme/interesting dynamics are highlighted with arrows and discussed in the main text.

COEFFICIENTS:	ESTIMATE	STANDARD ERROR	z-VALUE	$p (> z)$
(Intercept)	-8.2	2.8	-2.92	.003
Perceived audio similarity	1.8	0.9	2.02	.042
Algorithmic melodic similarity	0.04	0.02	1.99	.047
Algorithmic audio similarity	2.9	1.5	1.90	.057

Table 2 Logistic regression model after variable selection using stepwise AIC procedure. (For alternative models using different samples, variables, and/or Bayesian frameworks, see Supplementary Material Tables S1–5).

cases that highlight the following important dynamics of music copyright.

7.1 OBVIOUS MELODIC INFRINGEMENT SETTLED OUT OF COURT (*HARUMAKI GOHAN V. MORI*)

This case showing the highest similarity ratings by both humans and the PMI algorithm is instructive as a window into a major challenge of this research: selection bias. This case is part of the MCIR database, but unlike almost all other MCIR cases, this case was not accompanied by an official legal decision, but was instead settled out of court because the similarities were so striking. Indeed, the PMI algorithm confirms the obvious: the two melodies are 100% identical (though the lyrics are different; Figure 6).

Such settlements are actually common, but this case is unusual because the publishing company formally admitted infringement and fired Mori in a statement entitled “当社元提携作家による不正行為についてのお詫び” [“Apology for Infringement by Our Former Songwriter”]. In most similar cases, defendants employ “no fault” settlements where they pay the plaintiffs to resolve the case without any legal admission of infringement (e.g., *HaloSongs v. Sheeran*). The implications of this trend on

the present and future research are discussed below in the section entitled “the problem of reliable ground-truth legal decisions”.

7.2 OBVIOUS NON-INFRINGING SIMILARITIES IN TITLES (*FRANCESCATTI V. LADY GAGA*)

This case showed the lowest similarities for both humans and the PMI algorithm, and despite the similarities in the title (“Juda” vs. “Judas”), the rest of the lyrics were also rated as not very similar (mean 1.8 on a 1–5 scale; Figure 7). This represents the converse of the *Harumaki v. Gohan* case, where similar meritless infringement claims are very common, yet are usually dropped or settled out of court prior to a legal decision due to the expensive nature of a lawsuit and low probability of winning. Cases such as this that do make it all the way to a formal legal decision tend to have similar titles, perhaps because similarities in titles carry more weight or are easier to “sell” to a jury or media without requiring extensive musical analysis. Indeed, 10 of the 40 cases analyzed have similar or identical titles, though only 3 of these (*Three Boys v. Bolton*; *Grand Upright v. Warner*; and *Boosey v. Empire Music*) were judged as infringing.

Figure 6 A comparison of the chorus melodies of “Hachigatsu no Rainy” (top, plaintiff) and “M.A.K.E.” (bottom, defendant) shows they are 100% identical. Here and throughout this article, melodies (plaintiff top, defendant bottom) are transposed to the common tonic of C to enhance comparability, and identical notes are coloured in red.

Figure 7 A comparison of excerpts of the choruses of “Juda” (top, plaintiff) and “Judas” (bottom, defendant). This case showed some of the lowest similarity ratings for full audio, melody, and lyrics.

7.3 SUBCONSCIOUS MELODIC PLAGIARISM WITH MINIMAL PERCEIVED SIMILARITY (*BRIGHT TUNES MUSIC V. HARRISONGS*)

This case showed the most striking discrepancy between human and algorithmic ratings. Human participants listening to the full audio versions rate them as not very similar (<2 on a 1–5 scale), while the PMI algorithm identifies >50% identical notes in the melodic sequences (Figure 8). In this case, the melodic similarities were striking enough that a judge concluded that George Harrison’s “subconscious knew it already had worked in a song his conscious mind did not remember”, and found him liable for infringement to the tune of several million dollars.

7.4 SIMILARITY OF NON-ORIGINAL MUSICAL EXPRESSION (*VARGAS V. PFIZER*)

This case represents the final extreme of acoustic similarity in the absence of original melodic material. Instead, we find nearly identical rhythmic patterns played using nearly identical drum sounds (kick drum, snare, hi-hat) that nevertheless share no melodic similarities (as there is no melody, only rhythm; Figure 9). In this case, the rhythmic patterns were common enough that the judge granted summary judgment (dismissed the case without a full trial) on the grounds that the plaintiffs

had failed “to preclude the possibility of independent creation”.

7.5 UNRELIABLE GROUND TRUTH LEGAL DECISION (*WILLIAMS V. GAYE*)

Our final example comes not from the extremes of Figure 5 but from the middle to highlight the issue of unreliable ground truth decisions. This case ultimately found Pharrell Williams and Robin Thicke liable for infringing on Marvin Gaye’s “Got To Give It Up” in their Number 1 2013 hit “Blurred Lines”, awarding over \$5 million in damages. However, while the appeal process finally concluded with the 9th Circuit District Court affirming the decision, the correctness of the decision remains highly contested. The 9th Circuit District Court ruling included a dissenting opinion from one of the three judges, arguing that “the Gayes’ expert, musicologist Judith Finell, cherry-picked brief snippets to opine that a “constellation” of individually unprotectable elements in both pieces of music made them substantially similar.” Indeed, the “signature phrase” cited by Finell as being the smoking gun for similarity is short enough and shows minimal enough similarities (45%) as to not be significantly more similar than would be expected by chance from two random melodies composed from the same scales ($p = .2$; Figure 10).

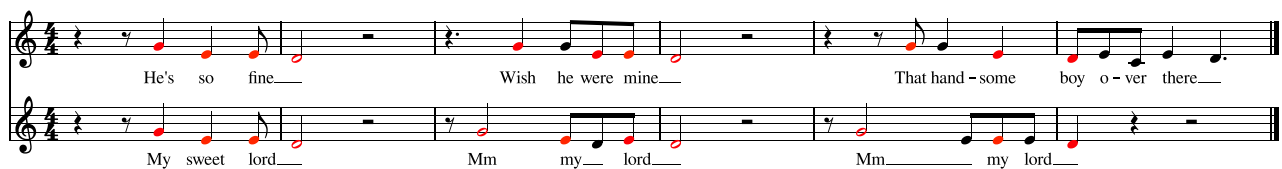


Figure 8 The openings of “He’s So Fine” (top, plaintiff) and “My Sweet Lord” (bottom, defendant) show high levels of melodic similarity. The judge concluded that they were “the very same song...with different words”. (Notation has been edited slightly from the version posted at the MCIR to match the audio recording.)

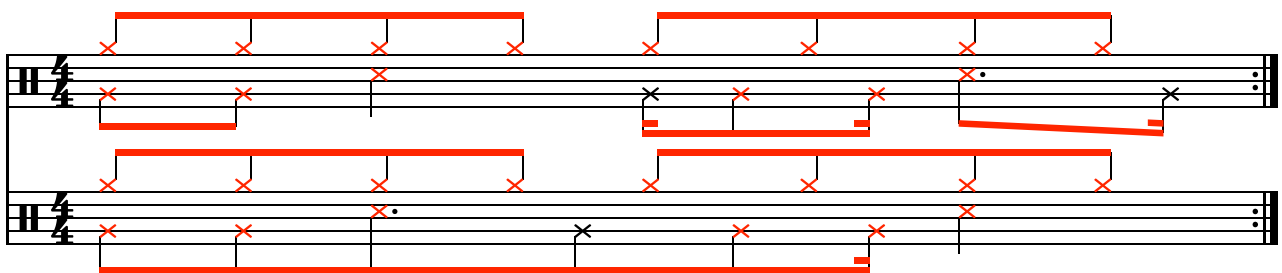


Figure 9 “Bust Dat Groove Without Ride” (top, plaintiff) and Pfizer’s background music in their commercial for “Celebrex” (bottom, defendant) use similar drum patterns, but this was not judged as infringing due to their non-original nature.

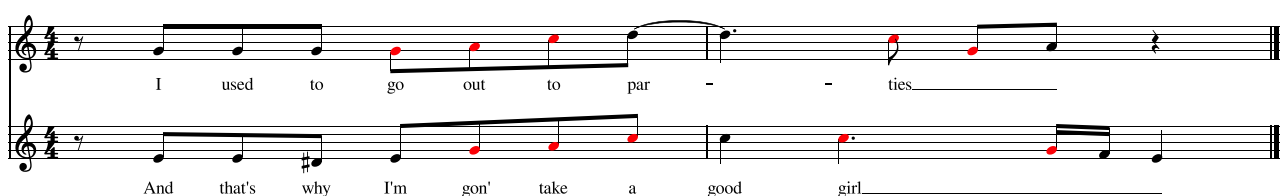


Figure 10 Comparison of the “signature phrases” of “Got to Give it Up” (top, plaintiff) and “Blurred Lines” (bottom, defendant) shows medium-low levels of similarity.

While the majority appeal decision concluded that the ruling was legally valid and “not reviewable after a full trial”, this does not necessarily imply that the jury’s decision was correct, but simply that it could not be legally over-ruled after the fact. Since this decision, other judicial opinions have come to opposite decisions, reversing jury decisions involving similar “constellations” of non-protectable elements (e.g., *Skidmore v. Zeppelin*; *Gray v. Perry*). Instead, the standard for such “constellations” appear to be correcting to the standard requiring “virtual identity” between the two works to establish infringement, as has been held in numerous earlier cases involving other copyrighted authorship like sculptural works (*Satava v. Lowry*, 2003) graphical user interfaces (*Apple Computer v. Microsoft*, 1994) and personal organizers (*Harper House v. Thomas Nelson*, 1989).

Our findings suggest that adding perceptual experiments or automated algorithms to this complicated process will not necessarily help. The PMI method does quantify a relatively low level of melodic similarity, while the full audio experiments suggest a somewhat higher but still not striking level of similarity. But many other cases with both higher and lower levels of melodic or acoustic similarity have been decided in both directions in the past, and the measures we use appear to be too coarse to offer conclusive evidence in such a polarizing case that has seen hundreds of musicians, musicologists, lawyers, and other interested parties weigh in on opposing sides.

8. DISCUSSION

Overall, using an expanded dataset of 40 cases including lyrics and other non-melodic features and more comprehensive signal detection and multiple logistic regression analyses, our current study confirmed one of the basic results of our preliminary study with 17 cases focused on melodic similarity (Yuan et al., 2020) but came to notably different conclusions in several important respects. First, the main point of consistency between the two studies is that, contrary to our predictions, listening to melody-only versions does not result in greater accuracy in matching past decisions. Instead, we found that listening to full audio versions actually predicted past decisions most accurately.

One important addition in the current study is that we have added analyses correcting for the fact that participants tend to err on the side of choosing non-infringement using two new analyses: a) signal detection analyses (ROC curves/AUC in Figure 2), and b) a multiple logistic regression model where the model intercept corrects for this bias (Tables 2 and S1–5). Notably, participants’ perceptions of similarity outperformed their estimates of infringement (Figure 2). Combined with the lower absolute accuracy of estimated infringement

(Figure 1), this suggests that our participants were not very good at making judgements on an absolute scale (because of the bias towards no infringement), but their judgements contain valuable information on a relative scale (especially their audio judgements, which give higher similarity ratings to infringing cases than non-infringing cases).

Using these new methods to correct for the bias toward “no infringement”, the current study found that human perceptual data, not automated algorithms, were the best predictors of past copyright decisions. This finding was consistent both for the new full sample of 40 cases and for the subset of 17 cases used in the preliminary study. This suggests our revised conclusion is not an artefact of our sampling methodology but rather reflects our improved analyses that better account for chance agreement and participant bias towards no infringement judgements. These improved methods also confirmed that judgments based on lyrics alone isolated from their musical context did not allow participants to predict past decisions beyond chance levels.

We observed moderate agreement between automated and perceptual judgments of music copyright infringement. The fact that PMI values were significantly correlated with perceptual similarity for both melody-only and full-audio provides validation for PMI as a perceptually relevant measure of melodic similarity and is consistent with the idea that melodic similarity plays a role in judgments of overall musical similarity (Allan et al., 2007). This is supported by the fact that the accuracy of the PMI method of 75% was comparable to previous studies (71% [Yuan et al., 2020], 80% [Savage et al., 2018]), even though the dataset no longer was limited to cases focused only on melodic similarity.

The lack of correlation between Musy’s audio similarity algorithm and perceptual similarity was surprising given that Musy’s algorithm has previously performed at or near the top in evaluations of general musical similarity (Flexer & Grill, 2016). This may be partly explained by Musy’s reliance on MFCCs to capture timbral and rhythmic similarity, not melodic similarity. Previous studies have shown that limited inter-rater reliability in judgments of musical similarity can limit the performance of automated algorithms (Flexer & Grill, 2016). Future analyses using supervised learning or other algorithms for capturing melodic similarity (Müllensiefen & Pendzich, 2009) may be able to improve performance, although the subjective nature of musical similarity will still place limits on the ability of any algorithm to match human judgments, especially for diverse music outside the Western tradition (Daikoku et al., 2022).

The fact that most participants judged “no infringement” for most cases even though slightly over half of cases were judged as infringement by the courts does not reflect random guessing, since our analyses showed that human perceptual judgments

were significantly above chance levels (Figure 2) and our preliminary study including randomized control pairs in the perceptual experiments showed accuracies of 100% for all three conditions (Yuan et al., 2020). Instead, the legal documents describing the decision process (provided at the MCIR links in Table 2) suggest that these mismatches are likely due to the fact that judges and juries also weighed non-acoustic contextual features in their decisions. Moreover, although the musicians and non-musicians performed similarly in our results (cf. Figure 1B), we cannot draw strong conclusions about the role of musical expertise due to the low number of musicians we recruited for our experiments. It would be worth including more participants with professional music knowledge or copyright knowledge in future studies.

The fact that human participants who made judgments only based on audio similarity without information about the historical or legal context systematically under-estimated levels of copyright infringement, but outperformed both algorithms when directly compared using the same methods, suggests that the complexities of copyright law are difficult to fully capture through measurement of similarity alone (whether this measurement is done by humans or algorithms). The relative emphasis on melody, lyrics, other musical aspects, and extra-musical legal factors changes from case to case, limiting the power of any single objective method.

9. THE PROBLEM OF RELIABLE GROUND-TRUTH LEGAL DECISIONS

A crucial limitation of our research design – shared with previous similar studies – is our reliance on previously adjudicated copyright cases for ground-truth data. Because plaintiffs and defendants are not likely to invest the substantial time and money needed for a lawsuit when they are very likely to lose, the most clear-cut cases of infringement and non-infringement are usually abandoned or settled out of court before reaching a final formal legal decision (cf. Sections 7.1 and 7.2 for the instructive exceptions of *Harumaki Gohan v. Mori* and *Francescatti v. Lady Gaga*). This creates a selection bias such that only the most ambiguous and controversial cases (e.g., *Williams v. Gaye*; cf. Section 7.5) make it into the MCIR, making it a fascinating legal resource but limiting its ability to provide a balanced and reliable ground-truth sample for empirical analyses like this. Such selection bias for particularly complex cases may help explain the relatively low overall levels of accuracy we found, as the “ground-truth” data may themselves have limited reliability (cf. Flexer & Grill, 2016, on how this problem of limited ground-truth reliability is a general challenge in MIR research). Future research on the quantitative analysis of copyright infringement

will need to develop creative new ways of assembling datasets that are less subject to selection bias (e.g., by leveraging existing cover song datasets, and/or experimentally manipulating songs to create different known levels of similarity/infringement) to achieve a more comprehensive understanding of perceptual and automated evaluations of music copyright infringement.

10. LIMITATIONS AND FUTURE DIRECTIONS

While our expanded analysis increased the size and diversity from our preliminary analysis (Yuan et al., 2020), it remains heavily biased toward US cases which make up the bulk of those listed at MCIR and dominate the global music industry. The generality of our findings is also limited by our pool of primarily Chinese participants without musical training. Expanding the breadth and depth of both copyright cases and participants would allow us to make stronger, more general conclusions in future, and help apply these to practical debates about music copyright in different countries throughout the world (Brauneis et al., 2022).

One challenge we attempted but failed to overcome was separating the roles of melody and accompaniment in the full audio. We conducted pilot experiments using a similar design after using Spleeter (Hennequin et al., 2021) to demix audio files into separate vocal melody and instrumental background versions, but the quality of demixing was highly variable, rendering the pilot results uninterpretable. Other experimental methods that we piloted but abandoned due to feasibility issues involved breaking down the MIDI melody version into separate rhythm only (unpitched drum sounds) and pitch only components. However, these results also proved too difficult to interpret – in particular, it was impossible to create a “pitch only” version of a melody without rhythm. The closest we could do was to create an iso-rhythmic version of the melody with all pitches given equal durations (i.e., all quarter notes). But the resulting melody sounded strange, as changing the rhythms resulted in disrupting the underlying musical metre, and we thus found the pilot results to be uninterpretable. Ultimately, we decided that focusing on full audio, melody-only, and lyrics-only versions would be the most important comparative data we could collect given the time and resources available for our experiments. Future studies, however, may wish to explore additional experimental paradigms such as these.

Another promising direction for future research involves improving the automated algorithm. We elected to use two existing theoretically motivated algorithms (PMI for MIDI, Musly for audio), but we suspect that with enough effort invested more accurate and/or flexible algorithms could be produced – particularly given

the rapid ongoing advances in machine learning and artificial intelligence in MIR (Choi et al., 2017; Agostinelli et al., 2023). Such algorithms might combine aspects of our PMI melodic algorithm, Musly timbral similarity algorithm, the “Plagiarism risk detector” being developed by Spotify (Pachet & Roy, 2020) and/or other algorithms that weight musical features based on their frequency in existing music corpuses or apply other machine learning approaches (Müllensiefen & Pendsch, 2009). We wish to emphasize, however, that our current results reinforce our previous conclusions that, while objective quantitative methods may help supplement traditional qualitative analysis (cf. Malandrino et al., 2022), “Trial by algorithm will never replace trial by jury, nor should it.” (Savage et al., 2018).

DATA ACCESSIBILITY STATEMENT

Musical stimuli, data and analysis code are available at <https://github.com/comp-music-lab/music-copyright-expanded>. The full experiments can be accessed at https://s2survey.net/music_copyright_fa for full-audio, https://s2survey.net/music_copyright_molo for melody-only and lyrics-only, and https://s2survey.net/music_copyright_voao for vocals-only and accompaniment-only. Detailed summaries and primary legal documents for all 40 cases are available at the Music Copyright Infringement Resource (MCIR; Cronin, 2018) and are linked in Table 1 and wherever else they appear in this article.

NOTES

- Note that much of the introduction and methods text is copied from our previous publication describing this preliminary study (Yuan et al., 2020). This publication was awarded “Best Application” at the 2020 International Society for Music Information Retrieval (ISMIR) conference (<https://www.ismir2020.net/awards/>), and thus the current expanded version of that study was invited to be submitted to *Transactions of the International Society for Music Information Retrieval*.
- Detailed summaries and primary legal documents for all 40 cases are available at the Music Copyright Infringement Resource (MCIR; Cronin, 2018) and are linked in Table 1 and wherever else they appear in this article.
- When we finalized our sample in October 2021, MCIR contained 301 cases, but only 40 of these met our criteria required to perform controlled experiments (see Methods for details). As of January 2023, it now contains 315 entries. MCIR aims to be a comprehensive source of all published legal decisions on the topic music copyright decisions issued within the US, and has recently begun expanding to also include more non-US copyright cases. However, published legal decisions only represent a tiny fraction of the number of copyright disputes, as the vast majority of copyright disputes are settled out of court without issuing a published legal decision.
- While the MCIR provides musical documentation for most of the works in the disputes it records, some works are unattainable, particularly those of obscure plaintiffs. For some of the contested works, particularly prior to the 1960s, there exists a musical score but no commercial recording, which makes a matched comparison of full audio and MIDI melody impossible. We did include cases with the reverse pattern (full audio

- available but no transcribed melody). In these cases, the first author transcribed the melody from the recording.
- While different cases vary in the degree they emphasize lyrical, melodic, and other similarities, it was not possible to reliably quantify or classify this relative contribution a priori, since most cases tended to rely on most factors to at least some degree. This information can be partially inferred from the results of our experimental analyses (e.g., cases with high accuracy in the melody-only condition are likely to focus on melodic similarity), but interpretation may also require additional contextual factors (e.g., questions of access described in the introduction).
 - All audio recordings and the full experiment can be accessed using the links in the “Data and Code Availability” section.
 - The full experiments can be accessed using the links in the “Data and Code Availability” section.
 - Thus “accuracy” does not incorporate any information regarding whether there is debate about the correctness of the jury/court’s decision (cf. Section 8, “The problem of reliable ground-truth legal decisions”).
 - Note that rhythms are not eliminated for the perceptual stimuli, only for the PMI calculation (cf. Savage & Atkinson, 2015 and the Discussion section of the present article regarding treatment of rhythm in the PMI method).
 - If we instead use previous cutoff values optimized on other (partially overlapping) datasets, we get the following results:
 - using PMI cutoff of 50% from Savage et al. (2018): 68% (27/40).
 - using PMI cutoff of 46.8% from Yuan et al. (2020): 70% (28/40).
 - If we instead use the previous cutoff value of 32.8% similarity optimized on our preliminary subset of 17 cases (Yuan et al., 2020), classification accuracy is 60% (24/40).

ADDITIONAL FILE

The additional file for this article can be found as follows:

- Supplementary Material.** Tables S1–S5. DOI: <https://doi.org/10.5334/tismir.151.s1>

ETHICS AND CONSENT

Permission to perform this study was granted by the Keio University Shonan Fujisawa Campus Research Ethics Committee (no. 211) to PES.

ACKNOWLEDGEMENTS

We thank our experiment participants and thank Sho Oishi and Quentin Atkinson for their contributions to our preliminary analysis (Yuan et al., 2020).

FUNDING INFORMATION

This work was supported by a Grant-In-Aid from the Japan Society for the Promotion of Science (#19KK0064) and by funding from the New Zealand Government, administered by the Royal Society Te Aparangi (Rutherford Discovery Fellowship 22-UOA-040 and Marsden Fast-Start Grant 22-UOA-052) to PES.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Conceptualization: PES, CC, DM, SF, YY; Methodology/ Analysis/Investigation/Visualization: YY, PES, DM; Project administration/Supervision/Funding acquisition: PES; Writing – original draft: YY, PES; Writing – revised draft: PES; Writing – review & editing: CC, DM, SF.

Yuchen Yuan and Patrick E. Savage made equal contribution.

AUTHOR AFFILIATIONS

Yuchen Yuan

Graduate School of Media and Governance, Keio University, Fujisawa, Japan

Charles Cronin

George Washington University Law School, Washington, D.C., USA

Daniel Müllensiefen  orcid.org/0000-0001-7297-1760

Department of Psychology, Goldsmiths, University of London, London, UK

Shinya Fujii  orcid.org/0000-0001-5003-2367

Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan

Patrick E. Savage  orcid.org/0000-0001-6996-7496

Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan; School of Psychology, University of Auckland, Auckland, New Zealand

REFERENCES

- Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., Sharifi, M., Zeghidour, N., & Frank, C.** (2023). MusicLM: Generating music from text. *ArXiv*. DOI: <https://doi.org/10.48550/arXiv.2301.11325>
- Allan, H., Müllensiefen, D., & Wiggins, G. A.** (2007). Methodological considerations in studies of musical similarity. *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, 473–478.
- Brauneis, R., Cronin, C., & Lim, D.** (2022, March 18). *Music copyright infringement: Global perspectives* [Virtual conference]. <https://mailchi.mp/uic.edu/8fjl6kpwdr-2683852>
- Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M.** (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4), 668–696. DOI: <https://doi.org/10.1109/JPROC.2008.916370>
- Cason, R. J. S., & Müllensiefen, D.** (2012). Singing from the same sheet: Computational melodic similarity measurement and copyright law. *International Review of Law, Computers & Technology*, 26(1), 25–36. DOI: <https://doi.org/10.1080/13600869.2012.646786>
- Choi, K., Fazekas, G., Sandler, M., & Cho, K.** (2017). Convolutional recurrent neural networks for music classification. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2392–2396. DOI: <https://doi.org/10.1109/ICASSP.2017.7952585>
- Cronin, C.** (2018). *Music Copyright Infringement Resource*. <http://mcir.usc.edu/>
- Daikoku, H., Ding, S., Benetos, E., Wood, A. L. C., Shimozone, T., Sanne, U. S., Fujii, S., & Savage, P. E.** (2022). Agreement among human and automated estimates of similarity in a global music sample. *Proceedings of the 2022 International Folk Music Analysis Workshop (FMA 2022)*, 26–32. DOI: <https://doi.org/10.5281/zenodo.7100288>
- Fishman, J. P.** (2018). Music as a matter of law. *Harvard Law Review*, 131(7), 1861–1923.
- Flexer, A., & Grill, T.** (2016). The problem of limited inter-rater agreement in modelling music similarity. *Journal of New Music Research*, 45(3), 239–251. DOI: <https://doi.org/10.1080/09298215.2016.1200631>
- Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M.** (2021). *Deezer Research: Spleeter*. <https://research.deezer.com/projects/spleeter.html>
- Lock, O., & O'Rorke, O.** (2022, July 29). “22 million songs a year, and only 12 notes available”: Pursuing a copyright claim in the music industry today. *Music:Ally*. <https://musically.com/2022/07/29/ed-sheeran-pursuing-a-copyright-claim-in-the-music-industry-today/>
- Lund, J.** (2011). An empirical examination of the lay listener test in music composition copyright infringement. *Virginia Sports and Entertainment Law Journal*, 11(1), 137–177. DOI: <https://doi.org/10.2139/ssrn.2030509>
- Malandrino, D., De Prisco, R., Ianulardo, M., & Zaccagnino, R.** (2022). An adaptive meta-heuristic for music plagiarism detection based on text similarity and clustering. *Data Mining and Knowledge Discovery*, 36(4), 1301–1334. DOI: <https://doi.org/10.1007/s10618-022-00835-2>
- Mandel, M., & Ellis, D.** (2005). Song-level features and support vector machines for music classification. *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, 594–599.
- Müllensiefen, D., & Pendsch, M.** (2009). Court decisions on music plagiarism and the predictive value of similarity algorithms. *Musicae Scientiae*, 13(1 Suppl.), 257–295. DOI: <https://doi.org/10.1177/102986490901300111>
- Needleman, S. B., & Wunsch, C. D.** (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48, 443–453. DOI: [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4)
- Pachet, F., & Roy, P.** (2020). *Plagiarism Risk Detector and Interface*. <https://data.epo.org/publication-server/document?iDocId=6403518&iFormat=0>

- Page, W.** (2021, November 3). Global value of music copyright jumps 18% to a record high of \$39.6bn in 2021: Could it have been even higher? *Tarzan Economics*. <https://tarzaneconomics.com/undercurrents/music-copyright-2021>
- Robine, M., Hanna, P., Ferraro, P., & Allali, J.** (2007). Adaptation of string matching algorithms for identification of near-duplicate music documents. *Workshop on Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection (PAN07)*, 37–43.
- Savage, P. E., & Atkinson, Q. D.** (2015). Automatic tune family identification by musical sequence alignment. In M. Müller & F. Wiering (Eds.), *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, 162–168.
- Savage, P. E., Cronin, C., Müllensiefen, D., & Atkinson, Q. D.** (2018). Quantitative evaluation of music copyright infringement. In A. Holzapfel & A. Pikrakis (Eds.), *Proceedings of the 8th International Workshop on Folk Music Analysis (FMA2018)*, 61–66.
- Schnitzer, D.** (2014). *Musly: An Open-Source Audio Music Similarity Library*. <https://www.musly.org>
- Schnitzer, D., Flexer, A., Schedl, M., & Widmer, G.** (2011). Using mutual proximity to improve content-based audio similarity. *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 79–84.
- Selfridge-Field, E.** (1998). Conceptual and representational issues in melodic comparison. *Melodic Similarity: Concepts, Procedures, and Applications [Computing in Musicology]*, 11, 3–64.
- Selfridge-Field, E.** (2018). Substantial musical similarity in sound and notation: Perspectives from digital musicology. *Colorado Technology Law Journal*, 16, 249–284.
- Serrà, J., Gómez, E., & Herrera, P.** (2010). Audio cover song identification and similarity: Background, approaches, evaluation, and beyond. In Z. W. Raś & A. A. Wiczorkowska (Eds.), *Advances in Music Information Retrieval (Studies in Computational Intelligence, vol.274)*, 307–332. Springer. DOI: https://doi.org/10.1007/978-3-642-11674-2_14
- Yesiler, F., Tralie, C., Correya, A., Silva, D. F., Tovstogan, P., Gómez, E., & Serra, X.** (2019). Da-Tacos: A dataset for cover song identification and understanding. *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, 327–334.
- Yuan, Y., Oishi, S., Cronin, C., Müllensiefen, D., Atkinson, Q. D., Fujii, S., & Savage, P. E.** (2020). Perceptual vs. automated judgments of music copyright infringement. *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR 2020)*, 23–29. DOI: <https://doi.org/10.31234/osf.io/tq7v5>

TO CITE THIS ARTICLE:

Yuan, Y., Cronin, C., Müllensiefen, D., Fujii, S., & Savage, P. E. (2023). Perceptual and automated estimates of infringement in 40 music copyright cases. *Transactions of the International Society for Music Information Retrieval*, 6(1), 117–134. DOI: <https://doi.org/10.5334/tismir.151>

Submitted: 03 October 2022 **Accepted:** 17 May 2023 **Published:** 05 October 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Transactions of the International Society for Music Information Retrieval is a peer-reviewed open access journal published by Ubiquity Press.