QUANTITATIVE EVALUATION OF MUSIC COPYRIGHT INFRINGEMENT

Patrick E. Savage
Keio University Shonan Fujisawa Campus
psavage@sfc.keio.ac.jp

Charles Cronin
USC Gould School of Law
ccronin@law.usc.edu

Daniel Müllensiefen
Goldsmiths, University of London
d.mullensiefen@gold.ac.uk

Quentin D. Atkinson
University of Auckland
q.atkinson@auckland.ac.nz

ABSTRACT

Unfounded music copyright lawsuits inhibit musical creativity and waste millions of taxpayer dollars annually. Our aim was to develop and test simple quantitative methods in order to supplement traditional qualitative musicological analyses and improve the efficiency and transparency of music copyright lawsuits. We adapted automatic sequence alignment algorithms from computational biology to create a "percent melodic identity" (PMI) method that was initially developed to measure the cultural evolution of folk music from different cultures. This method automatically quantifies and visualizes the percentage of identical pitch classes shared between two melodic sequences. We applied the PMI method to a corpus of 20 pairs of melodies that had been the subject of legal decisions and that had previously been analyzed using automatic methods. We found that the PMI method was able to accurately predict 80% (16/20) of previous decisions, with PMIs below 50% usually resulting in decisions of no infringement (11/13 cases), and PMIs above 50% usually resulting in decisions of infringement (5/7 cases). Importantly, each of the four outlying cases could be explained by contextual factors not related to melodic similarity (e.g., lyrics, access). Our results provide promise for improving music copyright evaluation by supplementing traditional qualitative components with quantitative methods and visualization tools that are simple enough to be useful to juries, judges, and other non-musicologists.

1. INTRODUCTION

Copyright serves the public good by encouraging the creation of innovative expression by granting authors a limited term during which they alone have the right to capitalize on their works. In music, unauthorized copying of melodies, lyrics, or other attributes of music has been legally prohibited since the 18th century. Initially, copyright law was designed to protect simply against wholesale copying of entire musical works (e.g., J. C. Bach's sonatas, the first case in which music was recognized as protected by copyright). Gradually, however, copyright case law broadened the scope of impermissible copying, such that even subconscious copying of parts of a melody could constitute infringement if such copying was substantially similar to protectable expression in the earlier work. Exactly how much and what types of musical copying qualify as "substantial" is a multimillion dollar question that is being actively and intensely debated (Cason & Müllensiefen, 2012; Cronin, 2015; Fishman, Forthcoming; Fruehwald, 1992). These debates have important practical consequences for all, as inappropriate music copyright lawsuits not only inhibit musical creativity, but also waste millions of taxpayer dollars annually that cover the adjudication of these disputes, not to mention the financial and temporal losses of individual defendants. One reason for this waste is that judicial evaluation of claims of musical similarity, on which these disputes are grounded, typically involves expert testimony by musicologists, who tend to use subjective, idiosyncratic, and time-consuming methods, tailored to the interests of the party that has retained them.

Unlike other arts (e.g., visual arts) where no single dimension has been given priority in copyright claims, music is unique in that one dimension – melody – has traditionally been the focus of copyright debates. For example, the court in King vs. Northern Music (1952) declared: "It is in the melody of the composition—or the arrangement of notes or tones that originality must be found. It is the arrangement or succession of musical notes, which are the finger prints of the composition, and establish its identity."

The rise of fields such as music information retrieval and music cognition have resulted in the development of automated melodic similarity algorithms and their application to musical copyright cases (Cason & Müllensiefen, 2012; Cronin, 1998; Mongeau & Sankoff, 1990; Müllensiefen & Frieler, 2004; Müllensiefen & Pendzich, 2009; Robine, Hanna, Ferraro, & Allali, 2007; Selfridge-Field, Forthcoming). For instance, Müllensiefen and Pendzich (2009) developed an algorithm 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 weighting formula for estimating perceptual salience. They found that optimizing this algorithm to a cut-off similarity value of 0.24 allowed them to accurately predict 90% (18/20 cases) of court decisions centered on questions of melodic similarity between 1976-2006.

While such algorithms have been somewhat successful, they have also been hard to translate into terms that are meaningful for non-scientists. Not only is it hard for jury members to interpret a salience function value of 0.24, but even this value is dependent on the makeup of the 14,000-pop song sample, and thus redoing these analyses using a different reference sample would result in different cutoff values. Juries, judges, and other interested parties would benefit from an intuitive measure of melodic similarity that depends only on the two melodies in question and can be easily visualized through simple notation.

The goal of this article is to propose and test a simple quantitative measure of musical similarity against a series of influential past decisions. Supplementing subjective qualitative interpretations with clear and intuitive quantitative guidelines by which to compare new cases with past cases should increase transparency and efficiency, reducing the chance of costly mistakes in the legal process and stemming the recent explosion of meritless claims that aim to force a quick payout from artists unwilling or unable to accept the risks of the current system.

2. THE PERCENT MELODIC IDENTITY (PMI) METHOD

We propose to adapt a "percent melodic identity" (PMI) method to musical copyright cases. This method is based on the automated sequence alignment and percent identity calculations used in molecular genetics to compare DNA and protein sequences (May, 2004). It was originally adapted to music in order to quantify the cultural evolution of English and Japanese folk song melodies in ways that could be meaningfully compared both with each other and with the evolution of other types of music from around the world (Savage & Atkinson, 2015). However, musical copyright represents an ideal application for this method, since copying of melodies with modification is simply another form of musical evolution. The PMI method is a general one that can also be applied to other types of folk and art music around the world (e.g., *gagaku*, Child ballads; Savage, 2017), justifying its inclusion in the *Folk Music Analysis* workshop despite its application to popular music.

The PMI method and other melodic sequence alignment algorithms are similar in principle to Judge Learned Hand's "comparative method" (Fishman, Forthcoming) for evaluating musical similarity. Like Hand, the PMI method begins by transposing two melodies transcribed in staff notation to a shared tonic, eliminating rhythmic information by giving all notes equal values[1], and then aligning and counting corresponding notes. However, while Hand's alignments were performed manually, the PMI method can take advantage of automated sequence alignment algorithms (Needleman & Wunsch, 1970) to eliminate subjectivity in alignment (although alignments can still be performed manually either from scratch or to correct errors in the automated alignment, as is also done in molecular genetics).

Automatic alignment requires penalties to be specified for opening or extending gaps in the alignment (represented by dashes in Fig. 1). Previously, we found that gap opening penalties (GOP) of 12 and a gap extension penalty (GEP) of 6 were the most successful in distinguishing whether two folk melodies shared ancestry (Savage & Atkinson, 2015; although further testing may be warranted in future to see whether these parameters are optimal for music copyright cases). Once the melodies are aligned, the number of identical notes (*ID*) are counted and divided by the average length of the two melodies (L_1 and L_2)[2] to calculate percent melodic identity (*PMI*, previously termed *PID* or "percent identity") as the percentage of identical notes shared between the two melodies, as follows:

$$PMI = 100\left(\frac{ID}{\frac{L_1+L_2}{2}}\right) \quad (1)$$

The PMI method can also be used to determine whether a given PMI value is statistically significant beyond what might be expected by two stylistically similar melodies that share similar scales. To do this, the PMI value for a given pair of sequence is compared against the distribution of 100 random PMI values given the same sequence lengths and compositions, as calculated by randomly reordering one of the sequences (Savage & Atkinson, 2015). Thus, an observed PMI value greater than 95% of randomly reshuffled values corresponds to a significant *P*-value of $<.05$.

Figure 1 shows an example of the PMI method using the famous case of Bright Tunes vs. Harrisongs. In this case, Judge Owen concluded that George Harrison had subconsciously plagiarized The Chiffons' *He's So Fine* because the melody of his song *My Sweet Lord* was "virtually identical". The PMI method is able to quantify this statement more precisely. For the opening three phrases shown in Figure 1, there are nine identical notes, while the average length of both melodies is 14, giving a PMI of 64%. When automatically aligning the full melodies of both songs, the PMI value drops slightly to 56% (27 identical notes, average length = 48 notes).

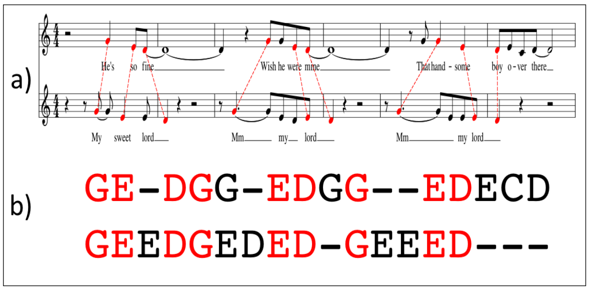

Figure 1. Comparison of the opening melodies of The Chiffons' *He's So Fine* (top) and George Harrison's *My Sweet Lord* (bottom) using ***a)*** standard staff notation, and ***b)*** the PMI sequence alignment method. In both cases, red represents aligned notes sharing identical scale degrees (joined with dashed lines in *a*). Dashes in *b* represent gaps inserted during the alignment process. PMI = 64% for these three phrases (56% for the full melodies). See Savage & Atkinson (2015) for details of how staff notation is converted into sequences of letters (including transposition to share a common tonic of C).

[1] Incorporating rhythmic information along with pitch information greatly increases computational complexity and does not appear to contribute substantial additional information (Cason & Müllensiefen, 2012; Cronin, 1998).

[2] This has been recommended as the most consistent denominator (May, 2008), but future investigation could explore whether other denominators may be more appropriate for music copyright. For instance, dividing by the length of the prior (plaintiff's) work may be more consistent with the principle that the shared content needs to constitute a substantial part of the original work.

3. MUSICAL COPYRIGHT INFRINGEMENT DATASET

For a ground-truth dataset to test the PMI method, we chose the set of 20 court decisions regarding melodic copyright infringement previously analyzed by Müllensiefen and Pendzich (2009). These decisions are a subset of the decisions available at the Music Copyright Infringement Resource (Cronin, 2018) selected by Müllensiefen and Pendzich because they contained clear rulings that were specifically focused on questions of melodic similarity (i.e., excluding cases focused on plagiarism of lyrics, unauthorized sampling of sound recordings, technicalities about the copyright registration process, etc.). This dataset seemed like an ideal starting point to test the PMI method against because the melodies had already been pre-selected and because automated similarity algorithms had already been used against them, providing a benchmark to compare the value of the PMI method. The list of cases is shown in Table 1, with the cases arranged by order of increasing PMI value.

4. RESULTS

4.1 Classification accuracy

Receiver operating curve (ROC) analysis using the area under the curve (AUC) measure confirms the intuitive impression from Table 1 that the optimal cutoff PMI value is 50% (AUC = 0.69). Using this cutoff, the PMI method was able to accurately classify 16 out of the 20 cases to match the court's decisions. For each of the four "failures", however, the following brief analyses show important non-melodic contextual factors that suggest that these exceptions were not due primarily to a failure of the melodic similarity algorithm but rather to the complex nature of musical copyright law (see Cronin, 2018 for further details on these and the other cases analyzed):

4.1.1 Grand Upright vs. Warner

There was no significant similarity between the melodies of Gilbert O'Sullivan's *Alone Again (Naturally)* and Biz Markie's *Alone Again* (PMI = 27%). However, there is obvious similarity in the lyrics, particularly in the title phrase "Alone again, naturally" used in both works. More importantly, Biz Markie uses an unauthorized sample of Gilbert O'Sullivan's piano accompaniment, and it appears that this was in fact the deciding factor in the case. Thus, this case may not have been appropriate for Müllensiefen and Pendzich to include (we have included it here for comparability).

4.1.2 Three Boys Music vs. Michael Bolton

This case is interesting because, although there is no significant melodic similarity between The Isley Brothers' *Love Is A Wonderful Thing* and Michael Bolton's song of the same name when taking the chorus as a whole (PMI = 36%), the opening phrase of each chorus uses not only the identical title lyrics but is also almost identical melodically (PMI = 86%; 5 out of 6 identical notes; Fig. 2). Thus, it appears that not only may similarities in the title/lyrics have influenced the jury's decision, but there may also be legitimate room for debate regarding how much of the melody should be included for purposes of melodic comparison and how long/complex a melody needs to be before it qualifies as original copyrightable expression.

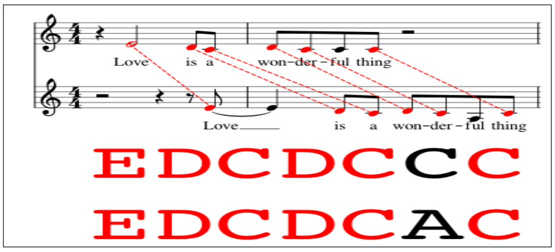

Figure 2. Comparison of the first phrase of the chorus of The Isley Brothers' *Love Is A Wonderful Thing* (top) and Michael Bolton's *Love Is A Wonderful Thing* (bottom). PMI = 86% for this phrase (but only 36% for the full choruses).

4.1.3 Selle vs. Gibb

The jury's verdict that the Bee Gees' *How Deep Is Your Love* infringed on Ronald Selle's *Let It End* was in fact consistent with the significant PMI value of 61% (Fig. 3). However, in this case the jury's verdict was overruled by the judge on appeal based on the fact that the Selle had not offered evidence to demonstrate that the Bee Gees had access to his work that would have allowed them to copy it. Such evidence is a legal requirement in addition to evidence of substantial similarity.

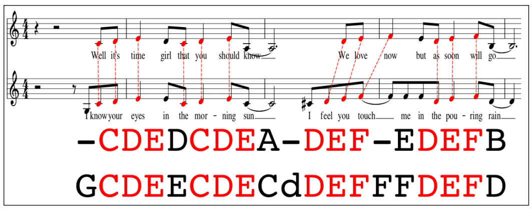

Figure 3. Comparison of the opening melodies of Ronald Selle's *Let It End* (top) and the Bee Gee's *How Deep Is Your Love* (bottom). PMI = 73% for these phrases (61% for the full chorus).

4.1.4 Fantasy vs. Fogerty

There was significant melodic similarity between John Fogerty's *Run Through The Jungle* and his *The Old Man Down The Road* (PMI = 67%) but a jury judged that the

No.	Case	Complaining work	Defending work	Decision	PMI
1	Suzane McKinley vs. Collin Raye	"I Think About You"	"I Think About You"	0	11%
2	Ferguson vs. N.B.C.	"Jeannie Michele"	"Theme 'A Time To Love'"	0	24%
3	*Grand Upright vs. Warner*	*"Alone Again (Naturally)"*	*"Alone Again"*	*1*	*27%*
4	Jean *et al.* vs. Bug Music	"Hand Clapping Song"	"My Love Is Your Love"	0	35%
5	*Three Boys Music vs. Michael Bolton*	*"Love Is A Wonderful Thing"*	*"Love Is A Wonderful Thing"*	*1*	*36%*
6	Cottrill vs. Spears	"What You See is What You Get"	"What U See is What U Get"	0	38%
7	Baxter vs. MCA	"Joy"	"Theme from 'E.T.'"	0	40%*
8	Intersong-USA vs. CBS	"Es"	"Hey"	0	40%**
9	Ellis vs. Diffie	"Lay Me Out By The Jukebox When I Die"	"Prop Me Up Beside The Jukebox (If I Die)"	0	40%
10	Granite Music vs. United Artists	"Tiny Bubbles"	"Hiding The Wine"	0	41%**
11	Repp vs. Lloyd-Webber	"Till You"	"Phantom Song"	0	45%**
12	McDonald vs. Multimedia Entertainment	"Proposed Theme Music 'Sally Jesse Raphael Show'"	"Theme Music 'Sally Jesse Raphael Show'"	0	46%
13	Benson vs. Coca- Cola	"Don't Cha Know"	"I'd Like To Buy The World A Coke"	0	49%*
14	Swirsky vs. Carey	"One of Those Love Songs"	"Thank God I Found You"	1	50%**
15	Bright Tunes Music vs. Harrisongs Music	"He's So Fine"	"My Sweet Lord"	1	56%**
16	Herald Square Music vs. Living Music	"Day By Day"	"Theme N.B.C.'s 'Today Show'"	1	56%**
17	*Selle vs. Gibb*	*"Let It End"*	*"How Deep Is Your Love"*	*0*	*61%***
18	*Fantasy vs. Fogerty*	*"Run Through The Jungle"*	*"The Old Man Down The Road"*	*0*	*67%**
19	Louis Gaste vs. Morris Kaiserman	"Pour Toi"	"Feelings"	1	73%**
20	Levine vs. McDonald's	"Life Is A Rock (But The Radio Rolled Me)"	"McDonald's Menu Song"	1	80%**

Table 1. The 20 music copyright infringement cases analyzed, ordered by increasing PMI (Percent Melodic Identity). See text for discussion of italicized exceptional cases. "0"=No infringement, "1"=Infringement. *$P<.05$, **$P<.01$.

two works were not musically substantially similar. The curious aspect of this case was that it involved a composer being accused by a recording company of plagiarizing his *own* work, the copyright to which he had assigned to the recording company. Given the substantial stylistic similarities expected among compositions by the same composer, it seems possible that the jury may have interpreted the judge's instructions regarding substantial similarity differently than they might for a case involving disputes between different composers. Furthermore, there was limited original expression in both melodies to begin with. Both are based predominantly on only two notes, so chance alone would already give a PMI of approximately 50%. In theory, this limited palette should be accommodated by the significance testing aspect of the PMI method, but – as discussed further below – this significance testing is complicated by other factors and cannot always be relied on.

4.2 Comparison with other algorithms

The best-performing algorithm tested by Müllensiefen and Pendzich (Müllensiefen & Pendzich, 2009) accurately predicted 90% (18/20) of these decisions. Their results were similar to our results using the PMI method, with the exception that Müllensiefen and Pendzich's algorithm resulted in *Three Boys Music vs. Michael Bolton* falling above their 0.24 optimal cutoff threshold, while *Fantasy vs. Fogerty* fell below this threshold.

Müllensiefen and Pendzich also tested other algorithms, including a "raw edit distance" algorithm that was more similar to the PMI method in that it was based purely on comparisons between disputed melodies without calibration against a database. The raw edit distance algorithm performed similarly to the PMI method, except that it failed to classify *Swirsky vs. Carey* as infringement[1].

However, as discussed above, it is not clear whether such differences in predicting court decisions truly imply that one melodic similarity algorithm is better. Indeed, the reverse may be true: Müllensiefen and Pendzich's algorithm may have overfit melodic similarity measures to match decisions that were affected by non-melodic factors such as lyrics or the identity of the composer. Future testing on a broader sample of cases should help determine whether there are substantive differences in the performance of these algorithms.

4.3 Statistical significance

The PMI method produced non-significant *P*-values in all cases where the PMI value was below 40%, and produced significant *P*-values in all cases where the PMI value was above 50%. PMI values between 40-50% gave mixed results, but generally produced significant *P*-values even though no infringement was found (5/7 cases). This suggests that the statistical significance measure is returning an inflated false positive rate. This is likely due to the fact that the comparison with completely random sequence is not a fair comparison, as even melodies that are completely unrelated will tend to share more pitch sequences than expected by chance alone due to universal regularities in melodic structure (e.g., tendencies for small, stepwise intervals and descending/arched melodic contours; Savage, Brown, Sakai, & Currie, 2015). We thus recommend caution in interpreting statistical significance of PMI values.

5. DISCUSSION AND FUTURE DIRECTIONS

We applied a simple and intuitive PMI (Percent Melodic Identity) method for measuring and visualizing melodic similarity to a ground-truth dataset of 20 court decisions on musical copyright. The PMI method performed similarly to existing, more complicated methods, accurately predicting 80% (16/20) of the decisions.

The major limitation of the current study is the limited sample of 20 cases and the fact that these cases include some complicating extra-musical factors. This makes it difficult to accurately evaluate automated melodic similarity algorithms against one another or conclusively determine whether they can usefully supplement future cases. In the decade since Müllensiefen and Pendzich compiled their sample, there have already been dozens of new decisions added to the Music Copyright Infringement Resource (including from countries such as China with different copyright regimes) and the number of active cases is increasing more rapidly than ever. In the future we plan to continue to expand and test the database to include these and many other cases from around the world. This broader sample will also allow us to address various technical issues such as the relative strengths of the similarity algorithms used, the effects of including rhythmic parameters, weighting different degrees of melodic similarity beyond simply identical or non-identical, etc. (Mongeau & Sankoff, 1990; Savage & Atkinson, 2015; Urbano, Lloréns, Morato, & Sánchez-Cuadrado, 2011; van Kranenburg, Volk, & Wiering, 2013). In particular, the cases discussed above highlighted the way similarity measurements can vary depending on the lengh of disputed melodic sections, and future studies may benefit by comparing different melodic lengths using both global and local alignment algorithms (van Kranenburg et al., 2013).

A broader issue is that the traditional reliance on melody as the key dimension by which to evaluate musical infringement may be changing along with the technology for making music (Cronin, 2015; Fishman, Fothcoming). This issue has been particularly actively debated recently following the jury verdict in *Pharrell Williams vs. Bridgeport Music* (currently under appeal) finding Williams and Robin Thicke liable for damages of over $5 million for infringing on Marvin Gaye's *Got To Give It Up* with their number-one hit *Blurred Lines* despite minimal melodic similarities. Specifically, the short "signature phrase" cited by expert musicologist Judith Finell as the primary melodic similarity, only gives a non-significant PMI value of 45% (5 identical notes out of 11; Fig. 4), while the PMI reduces to 19% when the full melodies are considered.

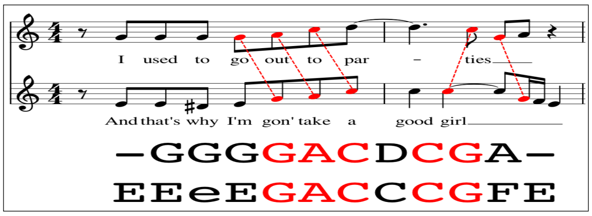

Figure 4. Comparison of the "signature phrase" from Marvin Gaye's *Got To Give It Up* (top) and Robin Thicke and Pharrell Williams' *Blurred Lines* (bottom). PMI = 45% for this phrase (19% for the full melody).

Hundreds of musicians, musicologists, and lawyers have weighed in on this decision, with some supporting the removal of arguably Eurocentric melodic notation from its dominant role in copyright law, while others fear the potentially stifling effect on creativity that vaguer and looser standards may cause (Cullins, 2018). Some have argued that we are already seeing a "Blurred Lines effect" (Fishman, Forthcoming) by which more dubious lawsuits

[1] In fact, *Swirsky vs. Carey* is the other case that may not have been appropriate to include in the database, as it was ultimately settled out of court without a final legal decision as to the question of infringement.

are being settled out of court due to fears that the old rules no longer apply. Given the changing norms in evaluating musical copyright infringement claims, and uncertainty about the relative weights given to the various factors going into past decisions, another important area for future work is to isolate the perceptual effects of melodic and extra-melodic similarities through controlled laboratory experiments (Lund, 2011; Müllensiefen & Frieler, 2004).

As even this small sample of cases shows, determinations of musical copyright infringement are too complex for it ever to be possible to predict outcomes perfectly through automated algorithms alone. Trial by algorithm will never replace trial by jury, nor should it. However, the automated, quantitative PMI method that we have presented is relatively accurate and easy for non-experts to understand and visualize. As such, we anticipate that it will help complement traditional qualitative analyses in future cases to create a more efficient, transparent, and just system for evaluating musical copyright infringement.

Acknowledgments: Funding support for this work was provided by a Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) scholarship to P.E.S. and a Rutherford Discovery Fellowship to Q.D.A. We thank the anonymous reviewers for comments on a previous draft.

6. REFERENCES

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.

Cronin, C. (1998). Concepts of melodic similarity in music-copyright infringement suits in melodic similarity: Concepts, procedures, and applications. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Computing and musicology 11* (pp. 187–209). Cambridge, MA: MIT Press.

Cronin, C. (2015). I hear America suing: Music copyright infringement in the era of electronic sound. *Hastings Law Journal*, *66*(5), 1187–1254.

Cronin, C. (2018). Music copyright infringement resource. Retrieved February 20, 2018, from http://mcir.usc.edu/

Cullins, A. (2018, December 28). Marvin Gaye's family gets boost from songwriters, musicologists in 'Blurred Lines' appeal. *Hollywood Reporter*. Retrieved from https://www.hollywoodreporter.com/thr-esq/marvin-gayes-family-gets-boost-songwriters-musicologists-blurred-lines-appeal-959465

Fishman, J. P. (Forthcoming). Music as a matter of law. *Harvard Law Review*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931091

Fruehwald, E. S. (1992). Copyright infringement of musical compositions: A systematic approach. *Akron Law Review*, *26*(1), 15–44.

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.

May, A. C. W. (2004). Percent sequence identity: The need to be explicit. *Structure*, *12*, 737–738.

Mongeau, M., & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities*, *24*, 161–175.

Müllensiefen, D., & Frieler, K. (2004). Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgments. *Computing in Musicology*, *13*(2004), 147–176.

Müllensiefen, D., & Pendzich, M. (2009). Court decisions on music plagiarism and the predictive value of similarity algorithms. *Musicae Scientiae*, *13*(1 Suppl), 257–295.

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.

Robine, M., Hanna, P., Ferraro, P., & Allali, J. (2007). Adaptation of string matching algorithms for identificaton of near-duplicate music documents. *Workshop on Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection (PAN07)*, 37–43.

Savage, P. E. (2017). 音楽の文化的進化を測る ――ブリティッシュ・アメリカンと日本の民謡・ポップス・古典音楽の事例を通して [Measuring the cultural evolution of music: With case studies of British-American and Japanese folk, art, and popular music]. PhD dissertation, Tokyo University of the Arts.

Savage, P. E., & Atkinson, Q. D. (2015). Automatic tune family identification by musical sequence alignment. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)* (pp. 162–168).

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8987–8992.

Selfridge-Field, E. (Forthcoming). Substantial musical similarity in sound and notation: Perspectives from digital musicology.

Urbano, J., Lloréns, J., Morato, J., & Sánchez-Cuadrado, S. (2011). Melodic similarity through shape similarity. In *Exploring Music Contents: 7th International Symposium, CMMR 2010* (pp. 338–355).

van Kranenburg, P., Volk, A., & Wiering, F. (2013). A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research*, *42*(1), 1–18.