Court decisions on music plagiarism and the predictive value of similarity algorithms

DANIEL MÜLLENSIEFEN* AND MARC PENDZICH**

* Department of Computing
Goldsmiths College, University of London
** Institute of Musicology
University of Hamburg

• **ABSTRACT**

Tune plagiarism in pop music is a common and often feverishly debated phenomenon which surely has to do with the vast amounts of money that individual melodies are able to generate in today's pop music business. The similarity between melodies is assumed to be a very important factor in a court's decision about whether a new tune is an illegitimate version of a pre-existing melody. Despite the wide-spread belief that there is a fixed and simple limit to the number of corresponding notes between two melodies, actual court decisions are based on far more complex considerations regarding the musical material.

This paper first sketches the legal framework and principal features of the legal processing of cases of alleged melodic plagiarism with a focus on US copyright law and discusses selected cases to highlight the corresponding legal practices. In the empirical part of this paper, we model court decisions for cases of alleged melodic plagiarism employing a number of similarity algorithms. As a ground truth dataset we use a collection of 20 publicly available cases from the last 30 years of US jurisdiction. We compare the performance of standard similarity algorithms (edit distance and n-gram similarity measures) to several new similarity algorithms that make use of statistical information about the prevalence of chains of pitch intervals in a large pop music database. Results indicate that these statistically informed algorithms generally outperform the comparison algorithms. In particular, algorithms based on Tversky's (1977) concept of similarity show a high performance of up to 90% of court decisions correctly predicted. We discuss the performance and structure of the algorithms in relation to a few interesting example cases and give an outlook on the potential and intricacies of our approach.

Keywords: plagiarism, music copyright, melodic similarity, similarity algorithms, Tversky's similarity model.

MOTIVATION

Melodic Plagiarism is a phenomenon that is hotly debated in the public arena every once in a while, particularly in the context of western pop music. It is not only in Germany that it has become an annual routine for the media to scrutinise the melodic originality of the songs of several hit-producing songwriters and to take a special look at the melody of entries for the Eurovision Song Contest. One of the aspects that make these cases so interesting to discuss is the connection between creativity and money in the form of royalties from author's rights. If one of the songs in a plagiarism case is frequently broadcast or forms part of well-selling recordings then the amount in dispute can easily be millions of euros.

Despite the huge public interest that plagiarism in pop music often raises, there is little research that is directly devoted to the matter: this is true of literature in both the pop music analysis and musical creativity fields, neither of which commonly discuss legal issues in detail.

A remarkable exception is Stan Soochers study "They Fought The Law" from 1999, in which he retraces a number of US-copyright cases of the last thirty years in a fascinating and very detailed way.

Another, more musicological exception is Charles Cronin's article on melodic similarity and copyright infringement (Cronin, 1998) where the author analyses American court decisions and different concepts of similarity as they were applied in legal arguments and by expert witnesses in the past. Although Cronin shows very elaborately how different understandings of melodic similarity were applied to the analysis of scores and melody transcriptions his article is not intended as a systematic and empirical study. With regard to this he does not include measurements of melodic similarity in any quantitative sense or relate algorithmic models to the similarity concepts he discusses. But Cronin's publication conveys a good impression of the many different ways melodic similarity might be related to court decisions on plagiarism.

The first aim of this study is to introduce the reader to several different categories to which cases of melody plagiarism can be assigned. The second is to explore how melodic similarity as measured by modern algorithms is related to court decisions in individual cases. To this end we measure the similarity of the melody pairs of twenty cases taken from a collection of court cases and evaluate the predictive power of the algorithmic measurements when compared to the court ruling. This evaluation follows a classifier evaluation paradigm common to many data mining tasks. We limit ourselves in this exploratory study in a number of ways: First, we choose to investigate only melodic plagiarism, which can be considered a reasonably homogeneous category of music plagiarism with the melody being the main item in dispute. This makes it possible to study this area using a specific subset of techniques from algorithmic

modelling. Other forms of music plagiarism are briefly discussed below. Second, we focus here primarily on US copyright law for two reasons. Firstly, because US copyright law can be sketched as one coherent system which is considerably different from European Authors' Right, and secondly because the evaluation sample that we are using comprises only cases that have been dealt with by US courts.

The empirical part of this paper is not intended as a comparative study of current similarity algorithms but as an exploration of how much insight standard similarity algorithms and statistically informed algorithms can shed on legal decisions concerning music plagiarism.

A (SHORT) HISTORY OF MUSIC PLAGIARISM AND LEGISLATION

During the course of the Renaissance era, authors and composers — who were formally considered rather as (sacred) artisans than (secular) artists — developed the concept of the creator of art works and increasingly considered their creations to be a product of their personal imagination and artistic choice. But along with the settlement of the concept of creative ideas as individual and personal creations came also the complementary case of the unauthorized use of creations: the *stolen* personal idea. Already in the 17th Century the assertion of ownership of foreign creations is denoted explicitly with the term **"plagiarism"** (Jörger, 1992, p. 29).

For example, L.v. Beethoven had the recurring problem during his time in Vienna, that other musicians hid themselves nearby his window and wrote down what they heard. Shortly after, the so-called composers presented in Vienna's society-salons the melodic themes by Beethoven as their own creations (Canisius, 1992, p. 187). Beethoven's problem with musical spies could possibly raise a smile on today's composers' faces. However, apart from the different circumstances and technical standards the basic problem — the unauthorised exploitation of a foreign intellectual work — seem to be essentially similar to today's copyright infringements.

In the absence of any legal protection, the only real options that Beethoven was left with to prevent his works from being plagiarised consisted in closing the window or playing the piano at a low volume. This legal situation changed in many European countries through the 19th century and by the beginning of the 20th century several (and improving) legislations concerning musical **intellectual property rights** were in place. Looking at today's situation at an international level, almost all countries have adopted one of the two fundamental and, with regard to their basic intent, completely contrary approaches of legislation, the continental-European author-based **Author's Right** and the Anglo-Saxon investor-based **Copyright Law**. Despite the conceptual differences at their origin, these two legal approaches have notably assimilated to each other over the last twenty years. The originally less-protecting Copyright approach with an often critisised more economic view on (musical) creations seems — *e.g.* in U.S. — to handle the new challenges of *digital* musical derivations (samples *etc.*) in a more practical way than the "good old" European Author's Right (Pendzich, 2004, p. 389f.). While the Author's Right was for a long

time considered to provide a tighter protection, now (since 1998) for the first time in the history of intellectual property, the (US)-Copyright law has overtaken the Authors' Right in terms of the temporal period for which copyrighted works *including* sound recordings are protected. In the US, the protection period for copyrighted sound recordings published before 1978 is extended to 95 years. As a consequence, for example the early sound recordings of Elvis Presley are still protected under US-Copyright Statute — but not any more according to EU-Author's Right which currently has a 50-year-limitation after publication of the recording[1].

To obtain an impression of the strong influence which the relevant legislation had on popular music history, it is worth having a look at the so-called classic era of rock music starting in the mid-1950s which was significantly moulded and determined by the **US-Copyright-Act of** 1909 (s. Pendzich, 2004, p. 110 ff.). To secure copyright under this law it was necessary to register the work "promptly" after publication. This was achieved by delivering two deposits containing a *copy* (sheet music, *not* a phonorecord[2]) to the **Copyright Office**. Additionally, it was required to file a copyright notice with the renowned circled "c" (©) on *each* copy of the work. A failure of a single aspect of this registration process resulted in the majority of cases into a complete loss of copyrights (and **royalties**) — and the musical work was immediately classified as being part of **public domain**. For US-authors it was absolutely necessary to have an accurate and detailed knowledge of how to secure the copyright for their works. This legal burden on the author stood in sharp contrast to the "automatic" author's right without any preconditions in the European tradition. But even if US-authors were partly aware of their rights, receiving their appropriate royalties wasn't always guaranteed. In recording studios especially young Rock'n'Roll musicians at the beginning of their careers like Chuck Berry and Carl Perkins often had only the choice between being recorded or not being recorded., while label managers or producers frequently signed the copyrights to themselves (Pendzich, 2004, p. 112). Legally, this bad practice was in general accordance with the Copy-right's **"work for hire"**-doctrine. The US-Copyright explicitly allowed for the possibility that **Copyright Owners** transferred their copyrights fully to others (with only very limited options to interfere with the artistic or commercial exploitation of the work later on).

As an amendment in this respect, the current US-Copyright Act of 1976, has implemented the concept of **initial authorship**. Transferred copyrights may be retransferred after 35 years — and the intricate copyright registration procedure is

(1) Note that there has been a recent initiative by the European Commission to extend the protection period of recordings beyond a 50 year period (see *e.g.* n.n. 2008).
(2) In the 1950s, phonorecords achieved more and more acceptance in the dispensation of justice: In S*hapiro, Bernstein & Co. v. Miracle Record Co.* the judge wrote in 1950: "It seems to me that production and sale of a phonograph record is fully as much a publication as production and sale of sheet music" (Patry, 1994, S. 419, Fn. 87).

no more required to secure copyright. Along with this amendment, the greater part of what one might consider today as undue hardship stemming from the US-Act of 1909 have been resolved to some degree in the meantime.

Copyright law focuses — in the explicit absence of the author's **moral rights** — mostly on the *use* of a copyrighted work. This encompasses the use of a musical work by **compulsory license** (in the form of a so-called **cover version** with a *limited adaption right* within the range of the **faire-use-doctrine**), as well as the use of the work as a **derivative work** or as a **phonorecord**.

In recent years the use of samples (= extract of a digital reproduction of another sound recording) has gained a very significant importance in modern pop music productions. The use of a sample requires at least compulsory license *and* **master use license** for the original recording.

In case of an (alleged) **infringement of copyright**, the **plaintiff** sues the **defendant** e.g. on damages in the way of the **civil lawsuit** at a **Federal District Court**. In case of appeal, there are 12 **circuits** with a **Court of Appeals.** Most of the lawsuits end at the appeal-level, but some of them are ruled in the last instance by the **Supreme Court** in Washington D.C.

This short and necessarily incomplete summary of legal parameters relevant for music compositions may have given insight of how the development of pop music history in general and the careers of individual composers, musicians, and music managers has been and continues to be deeply influenced and to be dependent of the legal framework within which commercial music is produced.

Conditions for melodic plagiarism

There are a number of "classical" conditions by which musical copyright can be infringed and that we need to discuss briefly. The most frequent cases include the (accused) *unauthorized* use of

- a copyrighted work *e.g.* as a short or extensive digital sample,
- a sound recording in a motion picture film or an advertisement,
- a copyrighted arrangement of a public-domain song,
- a too-similar rewrite of a musical work,
- similar lyrics with identical keywords *e.g.* in the chorus,
- the same work title with a partly-similar composition,
- or the claim of authorship for a third-party copyrighted work.

In addition to these main categories there are — especially in recent years — many more musical and legal conditions that have played a role in successful and unsuccessful claims concerning music plagiarism.

Considering pop music and leaving aside harmony, generally the melody is the most significant parameter of a musical composition. Many cases are connected directly or indirectly to the plagiarism of the *melody* of an original work of authorship, *i.e.* the use of a too similar sequence of notes, generally also including the rhythmic aspects of the melody.

261

In general, an original composition will be copyrightable containing only a small fraction of new musical ideas amongst the whole of its musical material: "The requisite level of creativity is extremely low; even a slight amount will suffice. […] Evaluation of creativity should be objective: the courts and the Copyright Office are not to judge the worth of creativity, but only its presence or absence" (Patry, 1994, p. 151). Building on this low threshold for originality and creativity, a typical strategy of a defendant will be to try to prove that the similar musical material, *i.e.* the matter of dispute, is pre-existent in an older work. If this is the case, then the plaintiff is not entitled to claim copyright for the musical material in question.

Another common defendant's strategy is to deny the knowledge of the plaintiff's initial composition. However, the proposition not have known the plaintiff's work is rarely accepted in court: "Once it appears that another has in fact used the copyright as the source of his production, he has invaded the author's rights. It is no excuse that in so doing his memory has played him a trick. In an indictment under Copyright Act, § 28 (Comp. St. § 9549), the excuse might be a defense, since the infringement would not be willful; but it is seldom that a tort, as this is, depends upon the purpose of the wrongdoer" (*Fred Fisher Music Co. vs. Dillingham*). All in all, the critical test is, whether the defendant had **access** to the older copyrighted work.

If the defendant cannot make it plausible that he infringed unwittingly or unconciously, the court will find a willful infringement for profit — and that entails further actions with criminal proceedings.

At a very general level, the important conditions to be met in a case of music plagiarism disputing the use of melodic material can be summarised as follows: The melodic material used in both works has be sufficiently similar, the melodic material of the plaintiff's work has to exhibit a minimal degree of originality and creativity, the plaintiff's work has to be (still) protected under copyright law and the defendant has to have had access to the plaintiff's work prior to the publication of his own composition. Unlike the fullfillment of necessary and sufficient conditions in scientific contexts, the judgement about these conditions in the legal world is rather a matter of complex opinions, debate, and even negotiation than a matter of simple measurement.

Method

The sample of cases

This study is based on a sample of US-copyright cases since 1970. All cases, including the opinions of the judges, are published. These documents are collected and provided on the internet *e.g.* by the "Columbia Law and UCLA Copyright Infringement Project" whose site[3] constituted a valuable source for this study. Almost all cases

(3) http://cip.law.ucla.edu/

listed on the project's website involve popular music (as opposed to western art music or folk music). This can be explained by the fact that all plagiarism suits are driven by a commercial motivation. The two fundamental requirements for making a lawsuit worthwhile pursuing are that

a) the defendant's work has to be commercially successful and

b) the plaintiff's work has still to be in copyright (*i.e.* composer of the musical work is not dead for more than 70 years).

By implication these requirements give an explanation for the predominance of popular music in plagiarism cases.

For this study we selected 20 cases from approximately 55 cases available online, spanning the years from 1970 to 2005. We chose only cases with a focus on melodic aspects of music copyright infringement. For these cases it was necessary to acquire the sound recordings in question, to analyse and edit the scores of the melodies according to our needs, to choose the relevant portions of the compositions and to convert the musical data to monophonic MIDI files. For a good proportion of the cases, the sound recordings, scores and (polyphonic) MIDI files were already available on the website of the copyright infringement project, but most of the melodic data had to be created (transcribed) from scratch or at least to be edited massively by the authors.

The written opinions of the judges were analysed in detail to make sure that they were based primarily on melodic parameters. Furthermore, we reduced the court decisions to only two categories: a) "pro plaintiff" denoting a positive instance of a "melodic plagiarism decision" or b) "contra plaintiff" meaning "no infringement". Some of the court decisions were not reducible in such a way because of the complexity of the case or because the verdict consisted rather of a collection of partial decisions regarding separate aspects of the case. By applying these criteria of minimal ambiguity and primacy of melodic material the collection we used for the algorithmic modelling was reduced to 20 cases. A tabular overview over all the cases is provided in appendix A.

In order to give a feel for the types of cases of melodic plagiarism that are taken to court (in the US) we will give short descriptions of few interesting cases from our test collection.

The earliest case of our collection, decided in 1976, is also probably the best-known copyright case in the history of pop music involving George Harrison's international post-Beatles hit "My Sweet Lord" from 1970 (*Bright Tunes vs. Harrisongs*). Already relatively similar regarding the more unspecific musical features, it was a single grace note in the main melody that confirmed the fact that it was undoubtedly take over from the initial work. This plaintiff's work, "He's So Fine" by The Chiffons, was played during 1963 by most British radio stations, so the access to the musical work was evident. The alleged fact that the defendant was unconscious of using the piece was not accepted as a reason to consider this a *fair use*. Therefore,

it was — according the court's decision — up to Harrison to pay for the publication of this unauthorized derivative work (s. Pendzich, 2004, p. 141ff.).

Another well-known example of proven plagiarism due to melodic similarity is case number 17 from our data sample, *Three Boys Music vs. Michael Bolton* (2000). Here, Bolton was sued and convicted for plagiarism of the Isley Brother piece "Love Is A Wonderful Thing" because of the musical and lyrical similarities of the chorus of his identically titled hit.

In contrast, *Tiny Bubbles versus Hiding The Wine* (case no. 2) of 1976 is one of the cases where plaintiff's claim was denied because the use of similar melodic material was limited to a few notes only and consisted of musical material with a very common use in pop music in general (*Granite Music vs. United Artists, 1976*).

Case number 3 belongs to the very well populated category of plagiarism suits where the plaintiff's claim was not successful due to a lack of similarity in the specific melodic material proposed to be taken over. The unpublished work "Jeannie Michele" was deemed not similar enough to John Williams' Soundtrack "A Time To Love" (*Ferguson vs. N.B.C.,* 1978). Leaving aside the fact that it could not be proved that the defendant actually had access to the plaintiff's unpublished work and looking at the case in retrospect, this appears to be an example where the plaintiff presumably had been misled by the alluring prospects of suing a hit composer and had overlooked the sparse musical commonalities between the two works. There are more candidates in this category of more or less obvious "trial-and-error"-cases where the plaintiff was probably aiming at reaching a convenient settlement with solvent defendants. One example (case no. 6) is the law suit versus the Coca-Cola Company's advertising song "I'd Like To Buy The World A Coke" alleging the relation to the work "Don't Cha Know" which bears little obvious similarity with Coca-Cola song (*Benson vs. Coca-Cola*, 1986). Or take the case *Cottrill vs. Spears* (no. 20, from 2003) where the music works had little more in common than their titles.

However, a clear reason to file a law suit had Ronald Selle with his musical work "Let It End", which bears a musical similarity at some level to the Bee-Gees-Hit "How Deep Is Your Love" (1984). In this case (no. 5) the jury's verdict was plagiarism, but the judge denied the access aspect of the case, and, as the circuit court regarded the failure of the proof of access to be of greater importance than obvious musical similarities, it was eventually decided in favour of the defendants, which has been critisised as being too harsh on the plaintiff (s. comment *Selle vs. Gibb*, 1984).

A different category covers cases where the two melodies in question exhibit very obvious similarities and the analyst wonders why a license for making use of an existing work wasn't acquired in the first place. This is the case *e.g.* for the note-by-note infringements from the "Theme For N.B.C.'s 'Today Show'" being *very* similar to Stephen Schwartz "Day By Day" (1978) (case no. 4). Similarly, one wonders how the literal copy of the melody from "Life Is A Rock" could not be regarded plagiarism when used in a McDonald's advertising campagne as the "Menue Song" (1990) (case

no. 9). Is it that big companies sometimes rather rely on their legal work-force than on musical common sense judgements?

By European standards a very astonishing and curious case is *Fantasy vs. Fogerty* (1994) (case no. 13), where the publisher of the successful Rock band Creedence Clearwater Revival sued the long-time main songwriter of the group, John Fogerty. Fogerty's song "The Old Man Down The Road" from his 1985 solo album Centerfield can indeed remind the listener of the CCR hit "Run Through The Jungle" from 1970. This situation created the rare situation that two rightholders of two works by the same composer engaged in law suit. However, in this case, the plaintiff's claim was denied.

This short overview over some of the cases from our data sample, spanning 30 years of US-Copyright jurisdiction, should have given an impression of the different types and motivations of music plagiarism cases and how largely varying degrees of similarities, causes of actions, and court decisions can be associated with this area.

Algorithms for measuring melodic similarity

Recent years have seen an enormous increase in the number of algorithms for measuring the similarity between monophonic melodies. The areas of application of these algorithm range from query-by-humming systems (McNab *et al.,* 1996; Dannenberg *et al.*, 2004) and score and incipit retrieval (Howard, 1998; Wiering *et al.*, 2004) in music information retrieval, to folk song research (Müllensiefen & Frieler, 2007) and ethnomusicology (Ahlbäck, 2007) and to music analysis (Nettheim, 1998) and psychological modelling (Müllensiefen, 2006). As diverse as their field of application is the algorithmic or mathematical construction of melodic similarity measure. Geometric measures (Ó'Maidín, 1998, Aloupis *et al.*, 2003), string matching techniques like edit distance (Mongeau & Sankoff, 1990; Crawford *et al.*, 1998), n-gram measures (Downie, 1999; Uitdenbogered, 2002), and hidden Markov models (Meek & Birmingham, 2002) from text retrieval and speech recognition were adapted for melodic information as was the Earth Mover's Distance algorithm (Typke *et al.*, 2007) from computer vision. Much seems to depend not only on the type of comparison algorithm used but also on the preprocessing of the melodic information in adequate terms. Here, the computation of statistical features (Eerola & Bregman, 2007) or the transformation into meaningful substructures (*e.g.* Weyde, 2004; Grachten *et al.*, 2004, Unal *et al.*, 2008) seems to warrant more robust results for different types of musical repertoires. Very few studies have been published that compare directly different algorithms for the same task and on the same dataset (*e.g.* Müllensiefen & Frieler, 2004) but some insight is available from three online comparison contests (MIREX, 2005, 2006, 2007) that have been held in the music information retrieval community (see *e.g.* Downie, 2006).

For the present purpose we decided to choose a few relatively simple and widely-used algorithms and to compare the performance of these standard algorithms in this plagiarism detection task to a class of algorithms that have experienced relatively little attention in melodic similarity studies so far and which make use of statistical information about the prevalence of musical structures in large music collections.

The "standard" algorithms

The edit distance

The edit distance or Levenshtein distance algorithm is not only one of the most standard algorithms for string comparisons in text processing and computational biology but it has also become almost a benchmark algorithm in music information retrieval (see *e.g.* Unal *et al.*, 2008, who compare their melodic fingerprinting algorithm to the performance of an edit distance algorithm). The edit distance is defined over two strings of symbols from the same alphabet as the minimum number of operations (insertions, deletions, substitutions) that is needed to transform one string into the other string.

For more formal definitions of the concept of edit distance and ways of implementing it the reader is referred to the general string matching literature (*e.g.* Gusfield, 1997) or the many contributions that apply it to melodies as strings of symbols (Mongeau & Sankoff, 1990; Crawford *et al.*, 1998). It suffices here to stress that the edit distance approximately models the notion of "how many notes have two melodies in common when we care about the order of the notes". This notion, though not explicitly a legal standard, is often more or less obscurely applied in expert witness reports that compare melodies for their similarities or overlap (see *e.g.* the diagrams in Cronin, 1998, pp. 195-96).

In principle edit distance can be applied to every suitable transformation of melodic data that results in a string of symbols, and in practice it has been applied to strings reflecting pitch, rhythmic or harmonic information as well as higher level abstraction derived from melodies (*e.g.* Grachten *et al.*, 2004). Instead of computing the global edit distance between all notes of two melodies, it would also be an interesting approach to identify the longest substring of notes that two melodies have in common (see *e.g.* Guo & Sigelmann, 2004; Lemström *et al.*, 2005). Focusing on the so-called *longest common subsequence* (lcs) would also help when two melodies a largely different lengths are to be compared. For this exploratory comparative study we limit ourselves to the application of the edit distance to pitch information. As our edit distance measure operates on the raw (*i.e.* untransformed) pitch information we like to refer it as Raw Edit Distance for the context of this study.

The other family of standard algorithms that we like to employ here as comparison measures are the so-called n-gram algorithms[4]. They have a widespread use in modern text retrieval they work on the basis of substrings of a specific length (*n*)

---

(4) We are aware of the fact that the term n-gram usually denotes a probability model where a sequence of n – 1 words or characters is used to predict the next one (= the nth one, see Jurafsky & Martin, 2000, p. 193). Nonetheless, for convenience we use the term n-gram in the remainder of this article to denote substrings (mainly of intervals) of length n, even if they are not associated with such a probabilistic model in the strict sense.

which are part of the two symbol strings to compare. In the following we call any string of consecutive symbols an *n-gram*. We call any specific string a *term τ*. Our use of these two technical terms in this paper corresponds to the technical terms *word token* and *word type* as they are known in computational linguistics (*e.g.* Baayen, 2001, p. 2). N-grams or word tokens are instances of, for example, melodic intervals or words that may or may not repeat in a melody or a text, such as "0-2" (note repeat and major 2nd down) or "twinkle". Terms or word types are distinct strings of musical symbols or letters. Another important concept in this context is the frequency by which a term *τ* occurs in a written document or a melody. We denote the frequency of term *τ* in melody *m* by $f_m(τ)$. To give an example, the following table lists the terms and term frequencies of all words and 2-grams of pitch intervals from the four opening bars of the well-known nursery rhyme *Twinkle, twinkle little star*. These four bars contain 4 verbal tokens (twinkle twinkle little star), and 3 distinct word types. Melodically these four bars comprise 14 pitches and hence 13 pitch intervals from which 12 pitch interval 2-grams can be derived. 9 distinct pitch interval 2-grams occur in the four opening bars.

Table 1

Verbal (word types) and melodic (pitch interval 2-grams) terms found in the four opening bars of Twinkle, twinkle little star with their corresponding frequencies. Pitch intervals are coded by the number of semitones they rise (positive integers) or fall (negative integers).

Verbal term *τ* (word type)	Frequency of verbal term *f(τ)*	Melodic term *τ* (pitch interval 2-gram)	Frequency of melodic term *f(τ)*
Twinkle	2	0, +7	1
little	1	+7, 0	1
star	1	0, +2	1
		+2, 0	1
		0, -2	3
		-2, -2	1
		-2, 0	2
		0, -1	1
		-1, 0	1

A few variants of n-gram algorithms have been proposed in the literature but the common underlying notion is that the number of terms and the frequency of the occurrence of each, when calculated for either one or both strings to be compared, is related to the similarity of the two. Again, the literature on string matching or n-gram comparison algorithms in general (*e.g.* Gusfield, 1997) and their application to music (Downie, 1999; Uitdenbogered, 2002) is not small and the interested reader is referred to it for detail and formal definitions.

Out of the several variants of n-gram measures we chose the *Ukkonen* measure and the *Sum Common* measures for the present study because they clearly differ in their approach to similarity measurement and both of them performed well in a previous evaluation (Müllensiefen & Frieler, 2004).

The Ukkonen measure

The Ukkonen measure counts the difference in the frequencies of all the n-grams occurring in both strings or in either string only and therefore reflects a notion of difference between the two strings to be compared. In the normalised version that scales the similarity values onto the interval from 0 to 1 the Ukkonen measure is defined for two melodies *s* and *t* as:

$$\sigma(s,t) = 1 - \frac{\sum_{\tau \in s_n \cup t_n} |f_s(\tau) - f_t(\tau)|}{|s| + |t| - 2(n-1)}$$

$f_s(\tau)$ and $f_t(\tau)$ are the frequencies of the term τ in melody *s* and *t* respectively, s_n and t_n designate the set of distinct terms in *s* and *t* and by |*s*| and |*t*| we denote the length of melodies *s* and *t*. The total number of n-grams in the two melodies are, respectively, $|s| - n+1$ and $|t| - n+1$.

The Sum Common measure

In contrast to the Ukkonen measure, which is concerned with differences in substring frequencies, the sum common measure sums the frequency of all n-grams occurring in both strings.

$$\sigma(s,t) = \frac{\sum_{\tau \in s_n \cap t_n} f_s(\tau) + f_t(\tau)}{|s| + |t| - 2(n-1)}$$

Both n-gram measures model the assumption that the number of common or different substrings (*i.e.* motives or melodic formulae in musical terms) or the frequency of these substrings is related to the overall similarity perception from the two strings in comparison. In contrast to the edit distance, the n-grams measure model the notion of "how many short motives have the two melodies in common if we don't care about the order of the substrings in the melodies".

For the sake of comparison we confine ourselves to two n-gram measures that operate on a pitch representation only; and because of the limited length of some of the melodies in our plagiarism database we limit ourselves to *n=3* which seemed to perform well previously (Müllensiefen & Frieler, 2004).

268

STATISTICALLY INFORMED SIMILARITY ALGORITHMS

The idea behind this rather heterogeneous family of similarity measures is to use information about the frequency or prevalence of individual melodic features in a particular style of music. The rationale here is that if two melodies share mainly very frequent features then their similarity is less significant as compared to two melodies that share mainly rare or unusual features. The features of a melody in this context can be any set of characteristics that can be computed without ambiguity from melodic data. This could include *e.g.* descriptors of melodic contour, rhythm, or implied harmony as well as pitches or intervals. For the scope of this study we limit ourselves to pitch intervals or rather n-grams of pitch intervals. In summary, the common aspects of the measures presented here are, firstly, the use of n-grams of pitch intervals as the melodic features or "raw" data, and secondly, the use of information about the frequency of the pitch interval n-grams in a large collection of music. We use a collection of 14,063 pop songs encoded as full polyphonic MIDI transcriptions of pop recordings from the 1950s to 2006 that was acquired in the context of a larger research project (the M⁴S project) at Goldsmiths College (see Müllensiefen, Wiggins, & Lewis, 2008, for details).

Apart from these two common aspects the conceptual idea and the mathematical construction of the here discussed statistically informed similarity measures differs to different degrees.

TF-IDF correlation

Pearson's correlation coefficient has been used in many similarity measurement situations where the two entities to be compared can be represented as vectors (see *e.g.* Kluge, 1974 or Steinbeck, 1982 for early applications to melodic data). It is also quite commonly applied as a similarity measure for text document retrieval where documents are conceptualised as vectors in a vector space model (Jurafsky & Martin, 2000, p. 647f).

In the context of this study we use the correlation coefficient with pitch interval n-grams that are weighted by their frequency of occurrence in the two melodies and their prevalence in the mentioned pop song collection.

The frequency of occurrence within a melody is generally referred to as *Term Frequency* (TF, see *e.g.* Manning & Schütze, 1999, p. 542). We define it here as a relative frequency, being a function of the term τ and the melody m with T indicating the number of distinct terms in m as follows:

$$TF(m, \tau) = \frac{f_m(\tau)}{\sum_{i=1}^{T} f_m(\tau_i)}$$

269

$f_m(\tau)$ denotes the frequency of term τ in melody m. This frequency is divided by the sum of the frequencies of all terms τ_i from $i=1,\ldots,$ T in m.

As said above, in the context of this study we use n-grams of pitch intervals as terms.

The prevalence of a term in a collection is measured by the *Inverted Document Frequency* (IDF, see *e.g.* Manning & Schütze, 1999, p. 543) which is defined as:

$$IDF_C(\tau) = \log\left(\frac{|c|}{|m:\tau \in m|}\right)$$

Here, $|C|$ denotes the total number of melodies in collection C and $|m:\tau \in m|$ is the number of melodies that contain term τ at least once.

A common way to combine TF and IDF weights for the terms of a given melody m with respect to a given collection C is by multiplication (see *e.g.* Manning & Schütze, 1999, p. 543; Jurafsky & Martin, 2000, p. 654). The rationale behind this combination of the two weighting schemes is to assign high weights to those terms that occur frequently in melody m but are not very common in the collection of melodies as a whole and, thus, can be regarded as being very specific for melody m:

$$TFIDF_{m,C}(\tau) = TF_m(\tau) \cdot IDF_C(\tau)$$

By inserting this weighting for the union of all the different terms in two melodies s and t in Pearson's correlation formula we obtain:

$$\sigma_C(s,t) = \frac{\sum_{\tau \in s_n \cup t_n} TFIDF_{s,C}(\tau) \cdot TFIDF_{t,C}(\tau)}{\sqrt{\sum_{\tau \in s_n \cup t_n}\left(TFIDF_{s,C}(\tau)\right)^2 \cdot \sum_{\tau \in s_n \cup t_n}\left(TFIDF_{s,C}(\tau)\right)^2}}$$

In summary, the TF-IDF correlation models the assumption terms that are frequent in both strings and infrequent in a large collection of melodies relate to a perception of high similarity. As an example for the application of a TFIDF-weighted similarity measure to a use case from music information retrieval see Uitdenbogerd (2002, p. 109).

TF-IDF common
In contrast to the just described correlation similarity the TF-IDF common measure acts on the intersection and not the union of all terms common to melodies s and t. The here proposed version is similar to the versions suggested by Uitdenbogered

(2002, p. 109) and Frieler & Müllensiefen (2007). The *TF* and *IDF* weights as defined above are used here again. The formula models again the notion that terms occurring frequently in both strings and being infrequent in the database relate to similarity perception.

$$\sigma_C(s,t) = \frac{\sum_{\tau \in s_n \cap t_n} \sqrt{IDF_C(\tau) \cdot \sqrt{TF_s(\tau) \cdot TF_t(\tau)}}}{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau)}$$

The square root of the product of the term frequencies of the same term τ from both melodies is multiplied by τ's inverted document frequency and the square root is taken from this product. The results are summed over all terms common to both melodies s and t and then normalised by the sum of the inverted document frequencies of all terms common to s and t.

Tversky's feature-based similarity

In a classic article Amos Tversky (1977) suggested that human similarity perceptions and judgements were based on the number of features two objects have in common and on the salience of these features. He proposed two families of similarity measures based in set theory one of which, the so-called "ratio model", appears to be compatible with conceptual idea and mathematical formulation of the TF-IDF measures just described. In its original formulation Tversky's ratio model has the form (1977, p. 333):

$$\sigma(s,t) = \frac{f(s_n \cap t_n)}{f(s_n \cap t_n) + \alpha f(s_n \setminus t_n) + \beta f(t_n \setminus s_n)}, \alpha, \beta \geq 0$$

Here $f(s_n \cap t_n)$ is a function that measures the salience or prominence of the features present in both melodies that are important for the notion of similarity. In analogy, $f(s_n \setminus t_n)$ and $f(t_n \setminus s_n)$ measure the salience of the features only present in s and t respectively. The choice of the weights α and β is crucial for the focus of the similarity comparisons. If $\alpha = 1$ and $\beta = 0$ then the salience of the features that s shares with t is only evaluated with respect to all features present in s. This choice of weights makes the similarity relation asymmetric and directional unless the two melodies are equal in their overall salience measure, *i.e.* $f(s_n) = f(t_n)$. Thus, for $\alpha \neq \beta$ generally $\sigma(s,t)$ is different from $\sigma(t,s)$ and σ is no longer a metric in the mathematical sense. If $\alpha = \beta = 1$ then the similarity model reduces to $f(s_n \cap t_n) / f(s_n \cup t_n)$. σ is a metric in this case which evaluates the salience of the shared features over the union of all features in both melodies.

271

The features of melodies that could be measured are potentially numerous and could include for example the melodic features employed by Eerola et al, (2006) or Eerola & Bregman (2007). For the scope of this exploratory study and for the sake of a comparison to the pitch-based n-gram and edit distance measures we limit ourselves here to substrings (n-grams) of intervals as features of the compared melodies.

According to Tversky, the salience function f is related to factors that can contribute to an overall perceptual salience of a stimulus like intensity, frequency, familiarity, good form, and informational content (Tversky, 1997, p. 332). In this study, we use the above described IDF weighting scheme as the salience function f. Since the IDF weights are derived from frequency counts in a large corpus of melodies they correspond to the notion of frequency in Tversky's concept and is also equivalent to the probabilistic concept of information content or self-information from information theory (Shannon, 1948). Information content is defined as the logarithm of the inverse of the probability of a specific outcome ω_τ of a random variable:

$$I(\omega_\tau) = \log \left(\frac{1}{P(\omega_\tau)} \right)$$

Taking a probabilistic view on the IDF weights one could ask for the probability that a melody m containing a specific term τ was drawn at random from a collection C and rewrite the IDF weighting probabilistically:

$$P(\omega_\tau) = \frac{|m : \tau \in m|}{|C|}$$

$$IDF_C(\tau) = \log \left(\frac{1}{P(\omega_\tau)} \right)$$

Out of the many possibilities to choose α and β we picked three pairs of these parameters which represent different point of view of how the similarity between two melodies conceptualised.

The Tversky.equal measure

For this measure we chose $\alpha = \beta = 1$. As said above this relates the intersection of terms (interval n-grams) between the two melodies to the union of terms. Inserting the IDF weights into Tversky's equation we obtain:

272

$$\sigma(s,t) = \frac{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau)}{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau) + \sum_{\tau \in s_n \backslash t_n} IDF_C(\tau) + \sum_{\tau \in t_n \backslash s_n} IDF_C(\tau)}$$

The Tversky.plaintiff.only and Tversky.defendant.only measure

The rationale for these measures is the common practice in the treatment of melodic plagiarism to evaluate the shared features of two melodies with respect to all features of only one of the melodies. In practice the argument can run in two ways: Either the melody of the defendant part is evaluated as to whether it contains significant original melodic material apart from the material shared with the plaintiff's melody. In this case the question to be answered is whether the original features eclipse the common features to a relevant degree.

But also the opposite strategy is found in legal arguments and expert witness reports: Does the melody of the defendant incorporate all of the (important) features of the pre-existing melody of the plaintiff? In this case only the plantiff's melody is considered as the reference context.

These two strategies can be modelled within the Tversky family of similarity measures just by setting one of the parameters α and β to *0* and the other one to *1*. If we denote the pre-existing melody of the plaintiff with s and the later published melody of the defendant with *t*, the two similarity measures are defined as:

$$\sigma_{\text{plaintiff.only}}(s,t) = \frac{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau)}{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau) + \sum_{\tau \in s_n \backslash t_n} IDF_C(\tau)}$$

$$\sigma_{defedant.only}(s,t) = \frac{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau)}{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau) + \sum_{\tau \in t_n \backslash s_n} IDF_C(\tau)}$$

The Tversky.weighted measure

Instead of choosing fixed values for α and β the parameters, for this measure we choose to determine their values dynamically for every pair of melodies. We make α and β dependent on the extent to which the shared n-grams cover the entire melodic material of melody *s* or *t* respectively. The values of α and β are therefore calculated as:

$$\alpha = \frac{\sum_{\tau \in s_n \cap t_n} TF_s(\tau)}{\sum_{\tau \in s_n} TF_s(\tau)} \qquad\qquad \beta = \frac{\sum_{\tau \in s_n \cap t_n} TF_t(\tau)}{\sum_{\tau \in t_n} TF_t(\tau)}$$

This choice of the two weighting parameters means that the more a melody is dominated by the shared interval n-grams the more weights it gains as a factor against which the shared n-grams are compared. α and β are necessarily from the range $0 \leq \alpha, \beta \leq 1$. This measure from the Tversky family makes not only use of the IDF weights but also exploits information about the term frequencies of the interval n-grams in both melodies.

Entropy weighting of statistically informed measures

Regarding the length of the terms or interval n-grams we adopted a flexible approach of averaging the similarity values of n-grams from length $n=1$ (*i.e.* simple pitch intervals) to length $n=4$. Each similarity value was weighted according to the entropy of the interval n-gram distribution of that particular length in the M[4]S pop song database (see above). The idea behind this weighting scheme is to award higher weights to similarity values computed on the basis of longer n-grams. This seems to make intuitive sense as longer exact, literal matches between two melodies can be considered to be contributing more to an overall similarity perception. The calculated weights were: 1 (1-grams, *i.e.* single interval), 1.66 (2-grams), 2.27 (3-grams), and 2.86 (4-grams). In addition to the information related to the occurrence of n-gram terms in a reference corpus reflected by the IDF weights this averaging over similarity measures for different n-gram lengths based on the entropy of the n-gram distribution is the second aspect that makes these measure statistically informed.

The comparison method

To compare the ability of these similarity measures to indicate legally relevant similarities between melodies we adopt a paradigm similar to those employed when comparing classification models (often simply called "classifiers") in binary classification tasks (see *e.g.* Hand, 1997): From each similarity measure we obtain a real-valued number from the range from 0 to 1. We have to compare this value against the binary court ruling of whether the two melodies in question constitute a case of plagiarism or not. A large number of performance measures has been defined to evaluate the relationship between a continuous classifier and a binary target variable. We make use of two of these established performance measures for comparing our collection of similarity measures: The first one is prediction accuracy at an optimal cutoff value. Here, we find an optimal threshold value for each similarity measure. All values above this threshold are then defined to indicate a sufficiently high similarity for a case of plagiarism whereas all values below the threshold are taken to indicate no plagiarism. The accuracy rate is then the number of correctly classified cases divided by the number of all cases in the sample (20 in this study). Often prediction accuracy is calculated with the additional information about the cost of the misclassification of a particular case. Unfortunately, we were not able to determine the actual costs of the cases in our database in terms of the legal fees and charges as well as the potential split of royalties in cases won by the plaintiff. Otherwise this would have been a wonderfully valid cost function to associate with prediction accuracy.

A significant disadvantage of accuracy as a performance measure is that the meaningfulness of its measurements is biased by the distribution between the two categories to be predicted. A prediction accuracy of, say 70%, does not sound bad at first glance but if, as in our case, only 7 out of 20 cases are instances of plagiarism then a relatively high number of correct predictions (65%) can be reached by just classifying all cases into one category (no plagiarism). Therefore, a good classifier should have a significantly better accuracy rate than a simple classification of all instances into the majority class.

To avoid this dependency of the distribution between classes of the target variable, a number of measures have become popular in signal detection and subsequently in psychophysics and medical classification to measure the performance of classifiers or predictors. A popular visualisation technique is the Receiver Operating Characteristic (ROC) curve (*e.g.* Swets, 1973) where the number of cases truly classified as positive is plotted against the number of cases falsely classified as positive for every value a classifier has produced on a given (test-)dataset (for illustration see ROC curves in section 3, *e.g.* Figure 3). For a good classifier a ROC curve rises steeply towards the upper left corner of the ROC diagram and then cuts across to the upper right corner. In contrast a chance classifier would generate a ROC curve close to the diagonal from left to right of the ROC space. A way to condense the performance of a classifier as depicted by the ROC curve to a single number is to measure the Area Under the Curve (AUC) which again is widely used in experimental psychology and medical diagnosis. The values of the AUC range between 0.5 (= chance performance, only the lower right half of the ROC space is covered) and 1 (= perfect performance, all of the ROC space is covered).

RESULTS

PERFORMANCE OF THE TESTED SIMILARITY MEASURES

Table 2 lists the classification accuracy and AUC results for the tested similarity measures.

The best performing measure is Tversky.plaintiff.only which classifies 18 out of 20 cases in accordance with the court decision (see also Figure 2 below). The similarity values fall mainly within the lower part of the value range from 0 to 1 and the optimal cut-off threshold for dividing pro-plaintiff and contra-plaintiff cases is 0.24. The two cases this measure gets wrong are case no. 8 and no. 12. We take a closer look at these cases in section 3.2 below. The best-performing standard measures are Raw Edit Distance and the Ukkonen measure which both classify 15 cases correctly and have threshold values of 0.46 and 0.29 respectively on the similarity scale.

The performance of two tested measures (the badly performing n-gram Sum Common and the best-performing Tversky.plaintiff.only) is visualised in Figures 1

Table 2

Classification accuracy and AUC values for the nine tested similarity measures

Similarity measures	Accuracy (# cases correct at optimal cutoff threshold)	AUC	Optimal cut-off on similarity scale
Raw Edit Distance	0.75 (15)	0.74	0.46
n-gram Sum Common	0.7 (14)	0.68	0.99
n-gram Ukkonen	0.75 (15)	0.76	0.29
TF-IDF correlation	0.85 (17)	0.85	0.13
TF-IDF common	0.65 (13)	0.58	-
Tversky.equal	0.85 (17)	0.85	0.29
Tversky.plaintiff.only	0.9 (18)	0.95	0.24
Tversky.defendant.only	0.7 (14)	0.64	0.24
Tversky.weighted	0.85 (17)	0.85	0.29

and 2. The plagiarism decision (1 = plagiarism, "∗" symbol; 0 = no plagiarism, "△" symbol) is plotted against the similarity values as given by the individual measure. A vertical line is placed at the optimal cut-off value at which all values equal or greater are classified as being instances of plagiarism.[5] It is quick to see that a measure with a low accuracy and a low AUC value like n-gram Sum Common produces a large region of overlap between the two categories. In contrast, the vertical line indicating the optimal similarity cutoff value of Tversky.plaintiff.only separates the two categories of court decisions quite well, *i.e.* the measure generally assigns higher similarity values to plagiarism cases and lower values to cases that were decided not to constitute plagiarism.

The same information can be summarised by ROC curve graphs which relate the magnitude of the similarity values to the number of true positives (plagiarism cases correctly identified by the algorithm) and false positives (cases of no plagiarism and incorrectly identified as plagiarism by the algorithm). The area right of the jagged line corresponds to the AUC values in Table 2. The more the line is bent towards the upper left corner, the higher the AUC value and the better the classification

1

(5) Since the different similarity measures vary in their coverage of the scale from 0 to 1, the optimal cut-off value for each similarity measure is chosen individually. This is chosen to maximize the accuracy of that measure's classification into plagiarism and non-plagiarism cases.
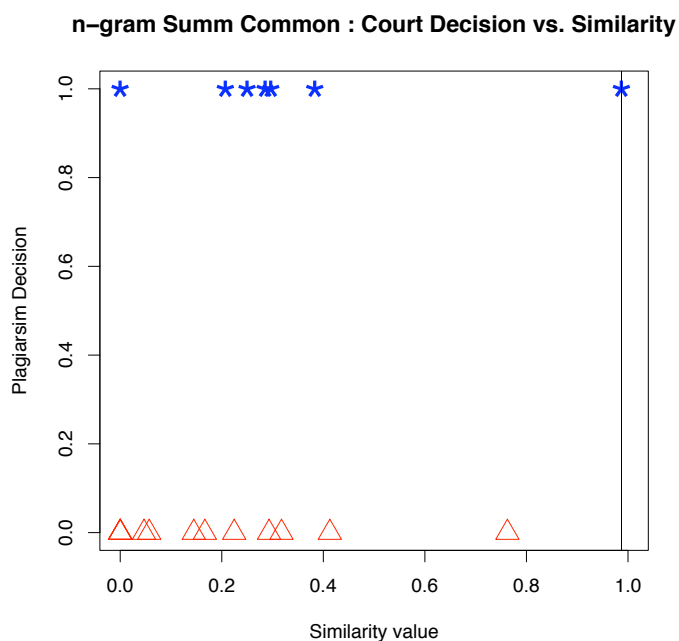
n–gram Summ Common : Court Decision vs. Similarity

Figure 1.

Classification performance of n-gram Sum Common, optimal cut-off at 0.99.

Tversky.plaintiff.only : Court Decision vs. Similarity
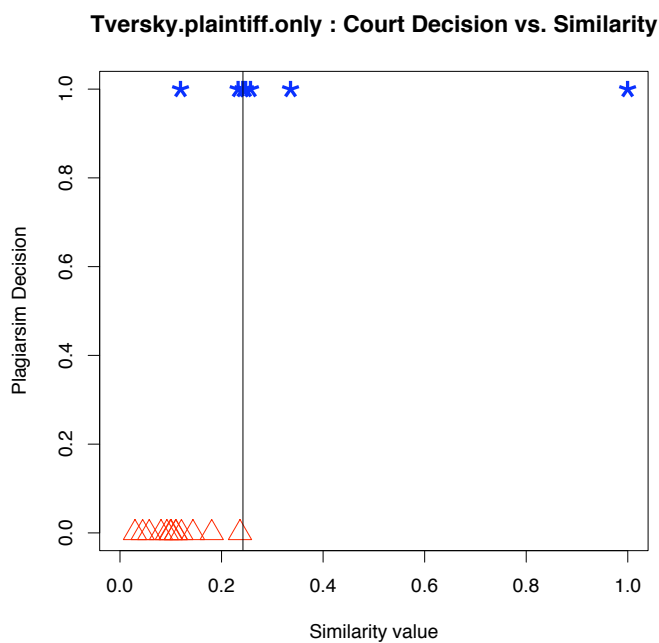
Figure 2.

Classification performance of Tversky.plaintiff.only, optimal cut-off at 0.24.

277

n–gram Summ Common : ROC curve

True positive rate

False positive rate
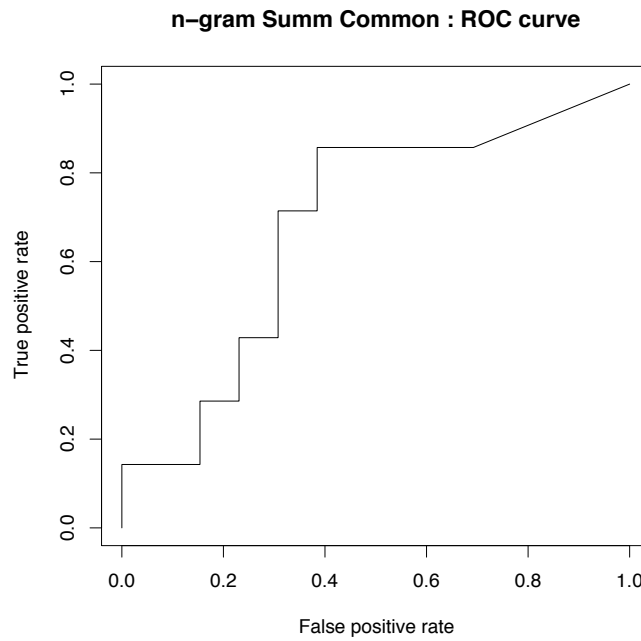
Figure 3.

ROC curve showing classification performance of n-gram Sum Common measure.

Tversky.plaintiff.only : ROC curve

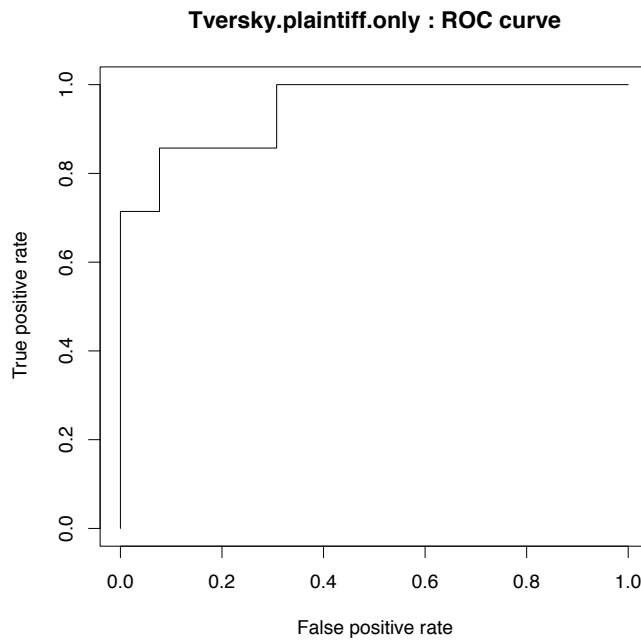True positive rate

False positive rate

Figure 4.

ROC curve showing classification performance of Tversky.plaintiff.only measure.

performance of the similarity measure. Figures 3 shows the ROC curve close to the imaginary diagonal generated from the performance of the n-gram Sum Common measure. Contrastingly, the area demarcated by the ROC curve resulting from Tversky.plaintiff.only in Figure 4 covers most of the area of the ROC space.

Two interesting general observations can be made from the numbers of Table 2: Firstly, the statistically informed similarity measures seem to perform generally better than the standard measures. This suggests that statistical information about the prevalence of interval n-grams in a large and suitable music corpus can indeed be a beneficial factor for similarity measures.

Secondly, the weak performance of the TF-IDF common measure and the Tversky.defendant.only suggests in addition that just using statistical information in a similarity measure does not guarantee a good performance of the measure. Instead, the details of the mathematical construction of the measure do matter and hence the question how melodic similarity is precisely modelled is of great importance.

As said before, the best performing measure is the Tversky.plaintiff.only measure which yields an AUC score of 0.95 and an accuracy of 0.9. This measure models the assumption that the salience of the shared features between two melodies are evaluated with respect to the salience of all features from the pre-existing melody. In particular it is interesting to note that this measure performed much better than Tversky. defendant.only, the other asymmetric similarity measure in our test collection.

We tested whether the superior performance of the Tversky.plaintiff.only measure was significant in comparison to the standard measures and the Tversky.defendant. only measure. We used a one-sided binomial test that tests whether the number of successfully classified cases (18 out of 20) of Tversky.plaintiff.only could have be happened by chance given the lower accuracy rate of the other measures. For Tversky. defendant.only (accuracy rate = 0.7) the test reached the usual significance level (p = 0.035) as was the case for the standard measures n-gram Sum Common (accuracy rate = 0.7, p = 0.035). However, the difference to the accuracy rate of Raw Edit Distance and the Ukkonen measure proved not to be significant at the 95% level (each: accuracy rate = 0.75, p = 0.091). Since we consider this study exploratory we renounced carrying out any further statistical exploration of the performance differences between similarity measures like bootstrapping or adjustment procedures for multiple testing. One has to bear in mind that the dataset we tested on is small in absolute numbers (n = 20) but at the same time covers already a large proportion of the entire population of court cases on melodic plagiarism from the US in recent times. Standard techniques that aim at providing a better or more robust estimate of a population parameter from a small sample might therefore generate misleading results.

A QUALITATIVE LOOK AT DISTINCTIVE CASES

Given the just discussed results having a closer look at three interesting and distinctive cases might illuminate how the different similarity measures work in

practice and they relate not just to the court's final verdict but also to the legal argument and circumstances of the particular case.

As outlined in section 2.1, case no. 5 was eventually decided as not being an instance of plagiarism despite some apparent similarities in the music. The melodies in question of the two songs are shown in Figures 5 and 6:

Figure 5.
Ronald Selle, "Let It End".

Figure 6.
Bee Gees, "How Deep Is Your Love".

The Raw Edit Distance gives a value of 0.585, above threshold, for these two melodies. Looking at the number of identical pitches falling into the same metrical positions it is not hard to understand why the jury judged these two melodies to be fairly similar. The final decision for this case was made on the basis of the fact that the defendant had almost certainly no access to the plaintiff's work. But the low similarity values from all of the statistically informed measures (between 0.000 and 0.218, all below plagiarism thresholds) might suggest another reason why this final verdict could be justified. The melodic parts shared by the two pieces are mainly diatonic scalar movements in minor and major seconds. The interval structures around the leaps of a fifth (bars 2, 4, 6, 8) of "Let It End" do not have direct correspondence in the Bee Gee's song. But the interval n-grams surrounding these big jumps, *e.g.* +2 +2 –7 +5 (bar 2, 4, in semitones), are much rarer in the pop melody repertoire than are up and down movements in seconds. However, high scores on statistically informed measures can only arise when rare terms (interval n-grams) are shared between the two melodies. If it is true that the Bee Gees had no access to the pre-existing song then the apparent similarities between the two melodies might be explained by a psychological mechanism based on the idea of

280

statistical learning and creation (see Wiggins *et al.*, 2009). It is, of course, perfectly possible that two composers write the same melody or have the same idea independent from each other (by the technical terms **independent creation** (McJohn, 2006, p. 38-40) or the German **zufällige Doppelschöpfung** (see the decisions by the German Federal Court of Justice, BGH, 1988a, 1988b) meaning an accidental or coincidental parallel creation by two individuals). But it is much more likely that they create similar melodies that overlap mainly by their more frequent and trivial melodic material from a common musical repertoire. In contract, creating similar melodies that share less frequent and highly specific melodic elements is much less likely. This perspective can only be modelled by similarity measures that take somehow advantage of the statistics of a musical style, like the above defined statistically informed measures.

Case no. 8 is Louis Gaste's "Pour Toi" vs. Morris Kaiserman's (Morris Albert) song "Feelings" (*Gaste vs. Kaiserman*, 1988). The Raw Edit Distance assigns these two melodic excerpts a value of 0.61 which is clearly above its cut-off threshold of 0.46.

Figure 7.
Morris Albert, "Feelings".

Figure 8.
Louis Gaste, "Pour Toi".

Looking at the scores in Figure 7 and 8 one can spot a number of shared intervals like the falling fifth in bars 1, 3 and 5 ("Feelings" only) and the step-wise motion just before this interval in bars 2 and 4. There are also clear differences like the different endings in bars 6 and 7 as well as the insignificant variations in the step-wise motions in bars 2 and 4 (exclusive use of whole tones and semitones in "Pour toi" versus scalar motion plus jump of a third in "Feelings"). But the Raw edit Distance apparently copes well with these differences by substituting and deleting notes.

In contrast, only some of the statistically informed measures judge this case correctly. The overall best performing measure Tversky.plaintiff.only does not predict this case correctly. This might be explained by the fact that at the beginning of bar 5 in "Pour toi" (the plaintiff's work) the falling fifth is replaced by a rising fourth. The

measures that evaluate the presence of this interval term (more) with respect to the other interval terms of Feelings (the defendant's work) are Tversky.equal, Tversky. defendant.only and Tversky.weighted. With these measures this term receives a greater importance since it appears three times in "Feelings" and because "Feelings" is shorter (18 notes altogether) and has, thus, overall fewer terms than "Pour toi" (23 notes). This case might therefore be seen as an example where the court evaluated the similarity between the two melodies in terms of the importance of a particular motive (*i.e.* falling fifth 1-gram and rising fifth plus falling fifth 2-gram respectively) in the defendant's work rather than its importance within the context of the pre-existing melody of the plaintiff. Regardless whether the value was below or above the threshold, all statistically informed measures gave a rather low value to this pair of melodies which is indicative of the fact that both melodies are constructed throughout on the basis of rather very common melodic intervals. Note repetitions, whole tones and semitones as well as the falling fifth are among the most frequent intervals in pop melodies. So, sharing these common intervals does not generate very high similarity values from the statistically informed measures.

The third case we like to look at in greater detail is case no. 12 in which the Irish singer-songwriter Raymond "Gilbert" O'Sullivan sued the New Yorker rapper Biz Markie over the title "Alone Again" (*Grand Upright vs. Warner*, 1991). From the commentary as given by the online copyright project it is not really clear in which similarities the judge based his decision. But apart from a sample of a piano ostinato the choruses of both songs exhibit a weak similarity as Figures 9 and 10 show.

Figure 9.
Biz Markie, "Alone Again".

Figure 10.
Gilbert 'O Sullivan, "Alone Again (Naturally)".

The characteristic end phrase of the chorus of the plaintiff's work (O'Sullivan) is composed of two half phrases the first one of which can be described as opening and falls into the upper register of the singer's voice while the second one closes the melodic line on the tonic and covers more the lower part of the vocal register. Biz Markie's chorus consists of two repetitions of a phrase that is similarly constructed

from two motives, the first being at a notably higher pitch than the second one. However, Markie's phrase uses a different intervallic content in various places and lacks the sweet resolution into a clear tonic on the last note of the second phrase.

From all the tested measures only the TF-IDF correlation measure generates a value above its threshold for this pair of melodies. N-gram Sum Common and n-gram Ukkonen even give a similarity value of 0 as no interval 3-gram is shared between the two short melodic excerpts.

This case seems to suggest that more features of melodies than just intervallic content can be taken into consideration. In this case it is probably the way the melody is constructed as a pair of phrases in two different registers in a call-and-response-type architecture. This high-level structural similarity is nothing that could be detected by any of measures tested in this study. In fact, it is uncertain whether any similarity measure that have been proposed in the literature would pick up on this type of structural similarity when the intervallic content differs to a degree like in this case no. 12. It would be interesting to try multiple viewpoint and pattern extraction approaches on this case that can combine pitch representations with temporal information and structural information at higher abstraction levels (*e.g.* Conklin & Anagnostopoulou, 2006; Lartillot & Toiviainen, 2007).

But of course it is only a guess, even though it seems to be a reasonable one, that the structural similarity between the two phrases was the decisive feature that lead to the plagiarism decision in court. Surely, the presence of this shared feature was very much assisted by the digital piano sample that Biz Markie had taken over from O'Sullivan's original recording and that shapes also the overall similarity perception of these two songs. Maybe the unauthorised use of the piano sample alone would have been sufficient for a plagiarism conviction but the approximate parallels of the chorus melodies made this inevitable. In addition to these musical circumstances one should not forget that court decisions are not automatic procedures (unlike algorithms) where a certain set of facts leads automatically to one specific court decision. Decisions are made and influenced by humans (the judge, the lawyers, expert witnesses, the public opinion) and thus court rulings can be as diverse as are the human beings involved in the verdict. In this case, Charles Cronin pays special attention to the personality of the judge Kevin Duffy in his online commentary (Comment on *Grand Upright vs. Warner*):

> Defendant's heart must have sunk upon learning that Judge Kevin Duffy would hear this case. The *Almanac of the Federal Judiciary* (2004) quotes lawyers who have appeared before Duffy in a mixed review that suggests he is an unpleasant and difficult judge to appear before: "He's mercurial. He can be a brute." One of the most often reversed judges in the Second Circuit, he was rebuked by a Circuit panel in 1996 for mistreatment of a lawyer appearing before him. In this opinion Duffy begins with a show-stopping biblical admonition — "thou shalt not steal" […]

It would be methodologically wrong to claim that with the help of these similarity measures we can identify a case of melodic plagiarism with a 90% confidence. Similarly, we would shy away from giving general advice that, say, an Edit Distance of > 0.46 or a Tversky.plaintiff.only value of > 0.24 for the similarity of two melodies necessarily signifies an instance of melodic plagiarism. But we are indeed surprised by the high accuracy rate (up to 90%) of some of the measures we tested. This is altogether even more surprising given the fact that apart from obtaining an optimal threshold we did not train our measures on this particular or any other collection of plagiarism cases. We only modelled assumptions about factors influencing melodic similarity by adapting similarity algorithms from the psychological and computational literature to work on musical pitch and pitch interval.

Tversky's original concept of similarity measurement hasn't found much repercussion in the literature concerned with melodic or musical similarity. One can only speculate about the reasons for this lack of involvement with Tversky's profound and prolific work on similarity perception. But possibly the requirement of valid salience functions as well as the notion that similarity judgements depend largely on the perspective of the judging individual and can thus be asymmetric (*i.e.*: $\sigma(s,t) \neq \sigma(t,s)$) make it less similar to the mathematical concept of a metric. Perhaps it is this property of Tversky's approach that might be regarded as less "elegant" or "straightforward" within an engineering context.

Contrasting this rather implicit notion, we obtain very good results from the Tversky measures that we implemented to work on pitch intervals only. This may indicate that a) Tversky's original concept for measuring the similarity between objects is applicable to melodies, b) using the IDF weights derived from a large pop song collection can be a useful salience function, and c) that evaluating melodic plagiarism is possibly best modelled with an asymmetric measure which (predominantly) uses the plaintiff's pre-existing work as context.

Apart from the particular success of the measures from the Tversky family it is also surprising that pitch content only was enough to obtain an acceptable classification accuracy. This applies also to the Raw Edit Distance and the Ukkonen measures (each 75% accuracy) and to the TF-IDF correlation measure (85% accuracy). This could mean that melodic plagiarism is decided about on the basis of pitch content in most cases. We showed above in the discussion of case no. 12 that exceptions to this general rule are possible and that other features like phrase architecture might occasionally come into play as well. This motivates of course the inclusion of features other than pitch or pitch interval into the framework of statistically informed similarity measures. Obvious candidates are the melodic transformations explored in Müllensiefen & Frieler (2004) such as relative duration, implied harmonic content or abstracted melodic contour. In principle, it is a

straightforward exercise to obtain IDF weights from the melodies of the here exploited pop song database of 14,000 MIDI files for those features (or rather sequences of features) as well. It would then be possible to build hybrid similarity measures which combine their knowledge about the prevalence of pitch interval n-grams with information about rhythmic sequences and the shape of melodic contours (*e.g.* Rizo *et al.*, 2008). Maybe then a case like no. 8 discussed above where a certain pitch interval (falling fifth) is combined with a certain duration ratio (short / very long) to form a characteristic motif would be detected as an important shared feature of the two melodies in this case. In addition to these rather basic melodic transformations it would surely be advantageous for the statistically informed measures to take higher level features such as general phrase architecture into account. But given the current small size of our database of plagiarism cases and given the performance of the pitch-based measures which approaches a ceiling effect, any optimisation would currently have only limited effect.

One of the primary future tasks of this project is therefore to broaden our database of court cases. This could be done by extending the list of US-American cases to cover most of the 20th century (going back until the last significant legal change, the revision of the US Copyright Act in 1909) or by including cases dealt with by *e.g.* English or German courts. But with all the necessity of extending the data basis for our algorithmic explorations, the collection of juridical verdicts and comments along with the actual music that was judged might prove to be the hardest part of the continuation of this study.

Despite all the encouraging results of this study and despite the options for constructing more comprehensive similarity models in the future we do not, of course, conclude or suggest that software algorithms could replace expert opinions as a means of judging or predicting cases of potential music plagiarism. As the discussion of case no. 12 showed the factors leading to a specific decision might be manifold and interacting with each other. Included are *e.g.* character traits of the human individuals that deal with a particular case in court. A comprehensive evaluation of this complex web of factors and dependencies can only truly be judged by a human expert with a experience in the domain and experience of the recent jurisdiction to which the case in question is to be subjected. Nonetheless, valid and reliable similarity algorithms can be highly useful to the human expert in order to highlight and quantify the features of the melodies in question that are relevant for a plagiarism investigation. In this respect similarity algorithms and statistical analysis can, firstly, inform precisely (*i.e.* numerically) about the extent to which features are shared between melodies, they can, secondly, inform whether the shared features possess the required degree of originality for copyrighted material and, thirdly, given a large database of relevant music, they can identify works pre-dating the plaintiff's composition that might contain similar or identical musical features. Especially the last point makes one advantage of computer technology applied in this domain very

clear: digital storage is effectively infinite and when used in combination with intelligent search and retrieval algorithms, potentially the entire history of a musical style can back up an investigation of a particular case of plagiarism. That way, the musical memory as one of the most important organs of an expert witness, which also is of paramount importance for his or her similarity judgements, can be extended by modern music information technology.

A further caveat for not taking the data mining exercise in this area too far is the nature of the ground truth data that we use, *i.e.* the court decisions on melodic plagiarism. Throughout this article we have assumed that the court decisions are correct and true, and of course legally speaking this is certainly the case until the decision is deemed to be wrong by a court of appeal. The reasons for a court decision to be wrong can be manifold as outlined in the previous paragraphs. Although very unlikely, the correction of particular decisions could in principle happen for many of the cases in our sample and given the low number of instances we base our statistical interference on, the revision of only a few cases could substantially alter the results presented above. We do not think that is going to happen with any discernible probability but just keeping the idea in mind that a plagiarism decision could indeed be wrong should prevent empirical researchers to draw too definite conclusions from the applications of algorithms to this type of music data from the real world.

Address for correspondence:
Daniel Müllensiefen
Department of Computing
Goldsmiths College, University of London
New Cross Road, New Cross
London SE14 6NW, UK
e-mail: d.mullensiefen@gold.ac.uk

• REFERENCES

Ahlbäck, S. (2007). Melodic similarity as a determinant of melody structure. *Musicæ Scientiæ*, Discussion Forum 4A, 235-80.

Aloupis, G., Fevens, T., Langerman, S., Matsui, T., Mesa, A., Rappaport, D., & Toussaint, G. (2003). Computing the similarity of two melodies. In *Proceedings of the 15th Canadian Conference on Computational Geometry*, Halifax, Canada: Dalhousie University, 81-4.

Baayen, H. (2001). *Word frequency distributions*. Dordrecht: Kluwer.

BGH [German Federal Court of Justice] (1988a). „Fantasy". In *Gewerblicher Rechtsschutz und Urheberrecht* (GRUR) 11/1988, 810-12.

BGH [German Federal Court of Justice] (1988b). „Ein bißchen Frieden". In *Gewerblicher Rechtsschutz und Urheberrecht* (GRUR) 11/1988, 812-16.

Canisius, C. (1992). *Beethoven. Sehnsucht und Unruhe in der Musik*. Aspekte zu Leben und Werk. Serie Musik Bd. 8297. München: Piper.

Columbia Law School & UCLA Law Copyright Infringement Project. Edited by Charles Cronin *et al*. Launched 2002. Update 5/2008. URL: http://cip.law.ucla.edu/

Conklin, D., & Anagnostopoulou, C. (2006). Segmental pattern discovery in music. *Informs Journal on Computing, 18(3)*, 285-93.

Crawford, T., Iliopoulos, C.S., & Raman, R. (1998). String-Matching techniques for musical similarity and melodic recognition. In Hewlett, W.B. & Selfridge-Field, E. (eds), *Melodic similarity. Concepts, procedures, and applications. Computing in Musicology* 11. Cambridge, MA: MIT Press, 73-100.

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

Dannenberg, R.B., Birmingham, W.P., Tzanetakis, G., Meek, C., Hu, N. & Pardo, B. (2004). The MUSART Testbed for Query-By-Humming Evaluation. *Computer Music Journal, 28(2)*, 34-48.

Downie, J.S. (1999). *Evaluating a simple approach to music information retrieval. Conceiving melodic n-grams as text*. PhD Thesis, University of Western Ontario, London (Ontario), Canada. URL: http://music-ir.org/~jdownie/downie_thesis.pdf.

Downie, J.S. (2006). The music information retrieval evaluation eXchange (MIREX). *D-Lib Magazine, 12(12)*. URL: http://www.dlib.org/dlib/december06/downie/12downie. html

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of Western and African folk melodies by Western and African listeners. *Psychology of Music, 34(3)*, 341-75.

Eerola, T. & Bregman, M. (2007). Melodic and contextual similarity of folk song phrases. *Musicæ Scientiæ, Discussion Forum* 4A, 211-33.

Fred Fisher vs. Dillingham, 298 F. 145 (1924). Source: *Columbia Law School & UCLA Law Copyright Infringement Project*. Edited by Charles Cronin *et al*. Launched 2002. Update 5/2008. URL: http://cip.law.ucla.edu/cases/case_fisherdill.html

Frieler, K. & Müllensiefen, D. (2007). Melodic similarity: The big picture and some new ideas. *Presentation given at the MIR Symposium and Workshop*, 20th Feb. 2007, University of Utrecht. URL: http://www.cs.uu.nl/research/projects/witchcraft/mir2007/

Grachten, M., Arcos, J. Ll., & López de Mántaras, R. (2004). Melodic similarity: Looking for a good abstraction level. *Proceedings of the 5th International Conference on Music Information Retrieval.* Pompeu Fabra University, Barcelona, Spain. URL: http://waits.cp.jku.at/~maarten/articles/GrachtenArcosMantaras-ISMIR04.pdf

Grand Upright vs. Warner. Comment. In *Columbia Law & UCLA Law School Copyright Infringement Project.* URL: http://cip.law.ucla.edu/cases/case_grandwarner.html

Gusfield, D. (1997). *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology.* Cambridge, UK: Cambridge University Press.

Hand D.J. (1997) *Construction and assessment of classification rules.* Chichester: Wiley.

Guo, A., & Sigelmann, H. (2004). Time-warped longest common subsequence algorithm for music retrieval. *Proceedings of the 5th International Conference on Music Information Retrieval.* Pompeu Fabra University, Barcelona, Spain. URL: http://ismir2004.ismir.net/proceedings/p049-page-258-paper113.pdf.

Howard, J. (1998). Strategies for sorting musical incipits. In Hewlett, W.B. & Selfridge-Field, E. (eds), *Melodic similarity. Concepts, procedures, and applications. Computing in Musicology* 11. Cambridge, MA: MIT Press, 119-28.

Jörger, T. (1992). *Das Plagiat in der Popularmusik…* Baden-Baden: Nomos.

Jurafsky, D. & Martin, J.H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* Upper Saddle River (NJ): Prentice Hall.

Kluge, R. (1974). *Faktorenanalytische Typenbestimmung an Volksliedmelodien.* Leipzig: VEB Deutscher Verlag für Musik.

Lartillot, O. & Toiviainen, P. (2007). Motivic matching strategies for automated pattern extraction. *Musicæ Scientiæ, Discussion Forum* 4A, 281-314.

Lemström, K., Navarro, G., & Pinzon, Y. (2005). Practical algorithms for transposition-invariant string-matching, *Journal of Discrete Algorithms, 3(2-4), Combinatorial Pattern Matching (CPM) Special Issue,* 267-92.

Manning, C. & Schütze, H. (1999). *Foundations of statistical natural language Processing.* Cambridge, MA: MIT Press.

McJohn, S. (2006). *Copyright: Examples and explanations.* New York: Aspen.

McNab, R.J., Smith, L.A., Witten, I.H., Henderson, C.L. & Cunningham, S.J. (1996). Towards the digital music library: Tune retrieval from acoustic input. In *Digital Libraries '96. Proceedings ACM Digital Libraries Conference,* 11-18.

Meek, C. & Birmingham, W. (2002). Johnny Can't Sing. A comprehensive error model for sung music queries. In Fingerhut, M. (ed), *Proceedings of the Third International Conference on Music Information Retrieval.* Paris: IRCAM, 124-32.

Mongeau, M. & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities* 24, 161-75.

Müllensiefen, D. (2006). "Exact measures of musical structure for predicting memory for melodies". In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (eds), *Proceedings of the 9th international conference of music perception and cognition* (ICMPC9), 782-90.

Müllensiefen, D. & Frieler, K. (2004). Cognitive Adequacy in the Measurement of Melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology* 13, 147-76.

Müllensiefen, D. & Frieler, K. (2007). Modelling expert's notion of melodic similarity. *Musicæ Scientiæ, Discussion Forum* 4A, 183-210.

Müllensiefen, D., Wiggins, G. & Lewis, D. (2008). High-level feature descriptors and corpus-based musicology: Techniques for modelling music cognition. In Schneider, A. (ed), *Hamburger Jahrbuch für Musikwissenschaft*, 23. Frankfurt: Peter Lang, 133-55.

Ó'Maidín, D. (1998). A geometrical algorithm for melodic difference in melodic similarity. In Hewlett, W.B. & Selfridge-Field, E. (eds), *Melodic similarity. Concepts, procedures, andaApplications. Computing in Musicology* 11. Cambridge, MA: MIT Press, 65-72.

Nettheim, N. (1998). Melodic pattern-detection using MuSearch in Schubert's Die schöne Müllerin. In Hewlett, W.B. & Selfridge-Field, E. (eds), *Melodic similarity. Concepts, procedures, and applications. Computing in Musicology* 11. Cambridge, MA: MIT Press, 159-68.

Patry, W. F. (1994). *Copyright law and practise.* Library of Congress Cataloging-in-Publication Data. Bureau of National Affairs, Inc. Washington, D. C.

Pendzich, M. (2004). *Von der Coverversion zum Hit-Recycling. Historische, ökonomische und rechtliche Aspekte eines zentralen Phänomens der Pop- und Rockmusik.* 2nd Ed. 2008. Münster: LIT.

Rizo, D., Lemström, K., & Iñesta, J.M. (2008). Tree structured and combined methods for comparing metered polyphonic music. In *Proceedings Computer Music Modeling and Retrieval 2008* (CMMR'08), Copenhagen, Denmark, 263-78.

Selle vs. Gibb. Comment. In *Columbia Law & UCLA Law School Copyright Infringement Project.* URL: http://cip.law.ucla.edu/cases/case_sellegibb.html

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, 379-423.

Soocher, S. (1999). *They fought the law. Rock music goes to court.* New York: Schirmer.

Steinbeck, W. (1982). *Struktur und Ähnlichkeit. Methoden automatisierter Melodieanalyse.* Kieler Schriften zur Musikwissenschaft XXV. Kassel: Bärenreiter.

Swets, J.A. (1973). The relative operating characteristic in psychology. *Science* 182, 990-1000.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-52.

Typke, R., Wiering, F. & Veltkamp, R. C. (2007). Transportation distances in human perception of melodic similarity. *Musicæ Scientiæ, Discussion Forum* 4A, 153-81.

Uitdenbogerd, A. L. (2002). *Music information retrieval technology.* PhD thesis, RMIT University Melbourne Victoria, Australia. URL: http://goanna.cs.rmit.edu.au/~alu/research/mythesis.pdf.

Unal, E., Chew, E., Georgiou, P. G., & Narayanan, S. S. (2008). Challenging uncertainty in query by humming systems: A fingerprinting approach. *IEEE Transactions on Audio Speech, and Language Processing, 16(2)*, 359-71.

Weyde, T. (2004). Modeling rhythmic motif structure with fuzzy logic and machine learning. In Hewlett, W.B. & Selfridge-Field, E. (eds), *Music query. Methods, models, and user studies. Computing in Musicology*, 13. Cambridge, MA: MIT Press, 35-50.

n.n. (2008). Urheberrecht. Cliff Richard soll auch als Rentner scheffeln. In: *Spiegel.* Issue 16/07/2008. URL: http://www.spiegel.de/kultur/musik/0,1518,566338,00.html

Wiering, F., Typke, R. & Veltkamp, R. (2004). Transportation distances and their application in music-notation retrieval. In Hewlett, W.B. & Selfridge-Field, E. (eds), *Music query. Methods, models, and user studies. Computing in Musicology* 13. Cambridge, MA: MIT Press, 113-28.

Wiggins, G.A., Pearce, M. & Müllensiefen, D. (2009). Computation al modelling of music cognition and musical creativity. In R. Dean (ed), *Oxford Handbook of Computer Music and Digital Sound Culture.* Oxford: University Press (in press).

Appendix A

List of plagiarism cases used for algorithm comparison including melodic similarity values as measured by three similarity measures (Tversky.plaintiff.only, Raw Edit Distance, Sum Common n-gram measure)

No.	Case	Song Plaintiff	Song Defendant	Decision	Similarity Tversky. plaintiff.only	Similarity Raw Edit Distance	Similarity Sum Common
1	Bright Tunes Music vs. Harrisongs Music 420 F. Supp. 177 (S.D.N.Y. 1976)	The Chiffons "He's So Fine"	George Harrison "My Sweet Lord"	1	0.242	0.547	0.297
2	Granite Music vs. United Artists 532 F.2d 718 (9th Cir. 1976)	Leon Pober "Tiny Bubbles"	Ernest Gold "Hiding the Wine"	0	0.144	0.375	0.225
3	Ferguson vs. N.B.C. 584 F.2d 111 (5th Cir. 1978)	Wilma Ferguson "Jeannie MiMichele" (unpublished)	John Williams "Theme 'A Time To Love'"	0	0.030	0.250	0.000
4	Herald Square Music vs. Living Music 205 U.S.P.Q. 1241 (1978); No. 77 Civ. 0008 (S.D.N.Y. Dec. 4, 1978)	Stephen Schwartz "Day By Day"	Ray Ellis "Theme N.B.C.'s 'Today Show'"	1	0.257	0.463	0.250
5	Selle vs. Gibb 741 F.2d 896 (7th Cir. 1984)	Ronald Selle "Let It End"	Bee Gees "How Deep Is Your Love"	0	0.058	0.585	0.0571
6	Benson vs. Coca-Cola 795 F.2d 973 (11th Cir. 1986)	John Benson "Don't Cha Know"	Coca-Cola Company "I'd Like To Buy The World A Coke"	0	0.045	0.349	0.293
7	Baxter vs. MCA, Inc. 812 F.2d 421 (9th Cir. 1987)	Leslie Baxter "Joy"	John Williams "Theme from 'E.T.'"	0	0.081	0.303	0.000
8	Louis Gaste vs. Morris Kaiserman 863 F.2d 1061 (2d Cir. 1988)	Louis Gaste "Pour Toi"	Morris Albert "Feelings"	1	0.233	0.609	0.286

9	Levine vs. McDonald's 735 F. Supp. 92 (S.D.N.Y. 1990)	Paul DiFranco and Norman Dolph "Life Is A Rock (But The Radio Rolled Me)"	McDonald's "Menu Song"	1	1.000	0.726	0.987
10	McDonald vs. Multimedia Entertainment, Inc. No. 90 Civ. 6356 (KC) (1991 U.S. Dist. LEXIS 10649 (S.D.N.Y. July 19, 1991)	Gerard McDonald: "Proposed Theme Music 'Sally Jesse Raphael Show'"	Dan Radlauer "Theme Music For 'Sally Jesse Raphael Show'"	0	0.110	0.139	0.047
11	Intersong-USA vs. CBS 757 F. Supp. 274 (1991)	Enrique Chia "Es"	Julio Iglesias "Hey"	0	0.236	0.296	0.413
12	Grand Upright vs. Warner 780 F. Supp. 182 (S.D.N.Y. 1991)	Raymond "Gilbert" O'Sullivan "Alone Again (Naturally)"	Biz Markie "Alone Again"	1	0.120	0.267	0.000
13	Fantasy vs. Fogerty 510 U.S. 517 (1994)	John Fogerty "Run Through The Jungle"	John Fogerty "The Old Man Down The Road"	0	0.181	0.617	0.763
14	Repp vs. Lloyd-Webber 132 F.3d 882 (2d Cir. 1997)	Ray Repp "Till You"	Andrew Lloyd-Webber "Phantom Song"	0	0.092	0.300	0.145
15	Ellis vs. Diffie 177 F.3d 503 (6th Cir. 1999)	Everett Ellis "Lay Me Out By The Jukebox When I Die"	Joe Diffie "Prop Me Up Beside The Jukebox (If I Die)"	0	0.101	0.410	0.167
16	Suzane McKinley vs. Collin Raye No.Civ.A.3:96-CV-2231-P 1999 U.S. Dist. LEXIS 1137 (N.D.Tex. Jan. 22, 1999)	Suzane McKinley "I Think About You"	Steve Seskin and Collin Raye "I Think About You"	0	0.110	0.424	0.318
17	Three Boys Music vs. Michael Bolton 212 F.3d 477 (9th Cir. 2000)	Isley Brothers "Love Is A Wonderful Thing"	Michael Bolton "Love Is A Wonderful Thing"	1	0.336	0.412	0.383

18	Swirsky vs. Carey 376 F. 3d 841 (9th Cir. 2004)	Seth Swirsky and Warryn Campbell "One of Those Love Songs"	Mariah Carey "Thank God I Found You"	1	0.248	0.351	0.207
19	Jean *et al.* vs. Bug Music	Leo Nocentelli, *et al.* "Hand Clapping Song"	Whitney Houston "My Love Is Your Love"	0	0.099	0.294	0.000
20	Cottrill vs. Spears	Michael Cottrill and Lawrence Wnukowski: "What You See is What You Get"	Britney Spears "What U See is What U Get"	0	0.121	0.250	0.000

292

• **Decisiones judiciales sobre el plagio musical y valor predictivo de los algoritmos de similitud**

El plagio de una melodía en la música pop es un fenómeno común y a menudo debatido febrilmente, que indudablemente está ligado a las enormes sumas de dinero que las melodías individuales pueden generar en el actual negocio de la música pop. La similitud entre melodías es un factor muy importante en la decisión de un tribunal que debe juzgar si un nuevo motivo es una versión ilegítima de una melodía existente. A pesar de la creencia extendida de que existe un límite establecido respecto al número de notas que se corresponden entre dos melodías, las decisiones judiciales actuales se basan en consideraciones mucho más complejas en cuanto al material musical.

Este artículo traza en primer lugar el cuadro legal y los principales elementos del procedimiento legal sobre presunto plagio melódico, concentrándose en la ley sobre derechos de autor de Estados Unidos, y examina un número de casos seleccionados para mostrar las prácticas judiciales correspondientes. En la parte empírica del artículo, presentamos las decisiones judiciales sobre casos de supuestos plagios melódicos utilizando varios algoritmos de similitud. Hemos utilizado como material de partida una colección de 20 casos accesibles al público, que corresponden a los 30 últimos años de jurisdicción en Estados Unidos. Comparamos el resultado de algunos algoritmos estándar de similitud (medidas *edit distance* y *n-gram*) con otros algoritmos de similitud nuevos que utilizan la información estadística sobre la prevalencia de cadenas de intervalos tonales en una amplia base de datos de música pop. Los resultados indican que estos últimos algoritmos fundados estadísticamente superan generalmente a los algoritmos de comparación. En particular, los algoritmos basados en el concepto de similitud de Tversky (1977) muestran un alto rendimiento de hasta el 90% en la predicción correcta de las decisiones judiciales. Examinamos la interpretación y la estructura de los algoritmos en relación con algunos ejemplos de casos interesantes, y damos una visión general del potencial y de la complejidad de nuestra aproximación.

• **Decisioni giudiziarie sul plagio musicale e il valore predittivo degli algoritmi di similarità**

Il plagio di un motivo nella musica pop è un fenomeno comune e spesso dibattuto fervidamente che ha a che vedere indubbiamente con la grande quantità di denaro che singole melodie possono generare nell'attuale business della musica pop. La similarità tra melodie è un fattore molto importante nella decisione di un tribunale che deve giudicare se un nuovo motivo sia una versione illegittima di una melodia pre-esistente. Nonostante la convinzione diffusa che vi sia un limite stabilito al numero di note corrispondenti tra due melodie, le decisioni dei tribunali oggigiorno si basano su considerazioni molto più complesse riguardanti il materiale musicale.

Questo saggio traccia innanzitutto il quadro normativo e i principali elementi dei procedimenti legali sul presunto plagio melodico concentrandosi sulla legge sui diritti d'autore degli Stati Uniti ed esaminando un numero di casi selezionati per

evidenziarne le pratiche legali corrispondenti. Nella parte empirica di questo lavoro creiamo un modello delle decisioni giudiziali di alcuni casi di presunto plagio utilizzando alcuni algoritmi di similarità. Come archivio di dati *ground truth* consideriamo un insieme di venti casi provenienti dalla giurisdizione americana degli ultimi trent'anni consultabili pubblicamente. Compariamo il risultato di alcuni algoritmi standard di similarità (l'*edit distance* e le misure di similarità n-grammi) con altri algoritmi di similarità nuovi che utilizzano le informazioni statistiche sulla prevalenza di catene di intervalli tonali in un'ampia banca dati di musica pop. I risultati indicano che questi ultimi algoritmi sono più efficaci rispetto ai primi. In particolare, gli algoritmi basati sul concetto di similarità di Tversky (1977) mostrano risultati elevati corrispondenti al 90% delle decisioni giudiziarie predette correttamente. Esaminiamo i risultati e la struttura degli algoritmi in relazione ad alcuni casi interessanti dando una visione generale del potenziale e delle difficoltà del nostro approccio.

• Décisions judiciaires sur le plagiat en musique et valeur prédictive des algorithmes de similarité

Le plagiat des airs dans la musique pop est un phénomène courant, et souvent fiévreusement débattu, qui est certainement lié aux énormes sommes d'argents générées par les mélodies individuelles dans le business que représente aujourd'hui cette musique. On peut supposer que la similarité entre mélodies est un facteur important dans la décision judiciaire de dire si oui ou non un nouvel air est une version illégitime d'une mélodie préexistante. En dépit de la croyance largement répandue selon laquelle il y a une limite figée, simple, au nombre de notes correspondantes entre deux mélodies, les décisions judiciaires actuelles sont basées sur des considérations beaucoup plus complexes concernant le matériel musical.

Cet article esquisse tout d'abord le cadre légal et les caractéristiques principales du processus judiciaire de cas de plagiat présumé de mélodie, avec une mise d'accent sur la loi de copyright américaine, et débat sur des cas sélectionnés afin de mettre en évidence des pratiques judiciaires équivalentes. Ensuite, dans la partie empirique de notre article, nous présentons des décisions judiciaires sur des cas de plagiat présumé de mélodie utilisant plusieurs algorithmes de similarité. Nous avons utilisé comme matériel de base une collection de 20 cas accessibles au public, datant des 30 dernières années de la juridiction des États-Unis. Nous comparons la performance d'algorithmes standards de similarité (mesures *edit distance* et *n-gram*) avec plusieurs algorithmes de similarité nouveaux, qui font usage de l'information statistique sur la prévalence de chaines d'intervalles de tonalités, dans une vaste base de données de musique pop. Les résultats indiquent que ces algorithmes statistiquement fondés surpassent généralement les algorithmes de comparaison. En particulier, les algorithmes basés sur le concept de similarité de Tversky (1977) montrent une performance élevée de plus de 90% de décisions judiciaires correctement prédites. Nous discutons de la performance et de la structure des algorithmes, en lien avec quelques exemples de cas intéressants, et donnons un aperçu du potentiel et des subtilités de notre approche.

• Gerichtsentscheidungen zu Musikplagiaten und die Vorhersagekraft von Ähnlichkeitsalgorithmen

Melodiendiebstahl wird in der Popmusik häufig und oft heftig diskutiert, was sicher mit den großen Geldsummen zusammenhängt, die einzelne Melodien im heutigen Popmusikgeschäft einspielen können. Die Ähnlichkeit zwischen Melodien wird dabei als ein sehr wichtiger Faktor in Gerichtsentscheidungen zu der Frage angesehen, ob neue Melodien illegitime Versionen von vorher existierenden Melodien sind. Trotz des weit verbreiteten Glaubens, dass es ein feststehendes und einfaches Limit bei der Anzahl der korrespondierenden Noten zwischen zwei Melodien gebe, basieren tatsächliche Gerichtsentscheidungen auf viel komplexeren Überlegungen, was das musikalische Material betrifft. Dieser Artikel skizziert zunächst den juristischen Hintergrund und die Hauptmerkmale von juristischen Fallbearbeitungen, bei denen es um Behauptungen von Melodieplagiaten ging. Dabei werden besonders das US-Kopierschutzgesetz untersucht und ausgewählte Fälle diskutiert, um die entsprechende Rechtspraxis hervorzuheben. Im empirischen Teil dieses Artikels modellieren wir unter Verwendung einiger Ähnlichkeitsalgorithmen Gerichtsentscheidungen zu Fällen behaupteter Melodieplagiate. Als Grunddaten verwenden wir eine Sammlung von 20 öffentlich einsehbaren Fällen aus den letzten 30 Jahren der US-Rechtssprechung. Wir vergleichen die Leistung von Standard-Ähnlichkeitsalgorithmen (editierte Distanz und n-gram Ähnlichkeitsmaße) mit verschiedenen neuen Ähnlichkeitsalgorithmen, die statistische Informationen zur Herkunft von Intervallketten in großen Popmusikdatenbanken verwenden. Die Ergebnisse zeigen, dass diese statistisch informierten Algorithmen generell besser als die Vergleichsalgorithmen sind. Besonders die Algorithmen, die auf Tverskys (1977) Ähnlichkeitskonzept beruhen, zeigen eine hohe Leistung in der korrekten Vorhersage von bis zu 90% der tatsächlichen Gerichtsentscheidungen. Wir diskutieren die Leistung und Struktur dieser Algorithmen in Bezug auf einige interessante Beispiele und bieten einen Ausblick auf Potentiale und Schwierigkeiten unserer Herangehensweise.