(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 742 433 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
25.11.2020 Bulletin 2020/48

(51) Int Cl.:
G10H 1/00 *(2006.01)*

(21) Application number: **19176232.7**

(22) Date of filing: **23.05.2019**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(71) Applicant: **Spotify AB**
111 53 Stockholm (SE)

(72) Inventors:
- **PACHET, François**
  111 53 Stockholm (SE)
- **ROY, Pierre**
  111 53 Stockholm (SE)

(74) Representative: **Kransell & Wennborg KB**
P.O. Box 27834
115 93 Stockholm (SE)

Remarks:
Amended claims in accordance with Rule 137(2) EPC.

(54) PLAGIARISM RISK DETECTOR AND INTERFACE

(57) Methods, systems and computer program products are provided for testing a lead sheet for plagiarism. A test lead sheet having a plurality of passages is received at a plagiarism detector. A set of annotations describing a level of plagiarism of a plurality of elements (e.g., chord sequence, subsequences, melodic fragments (i.e., notes), rhythm, harmony, etc.) of the test lead sheet in relation to preexisting lead sheets are generated and output via an output device.

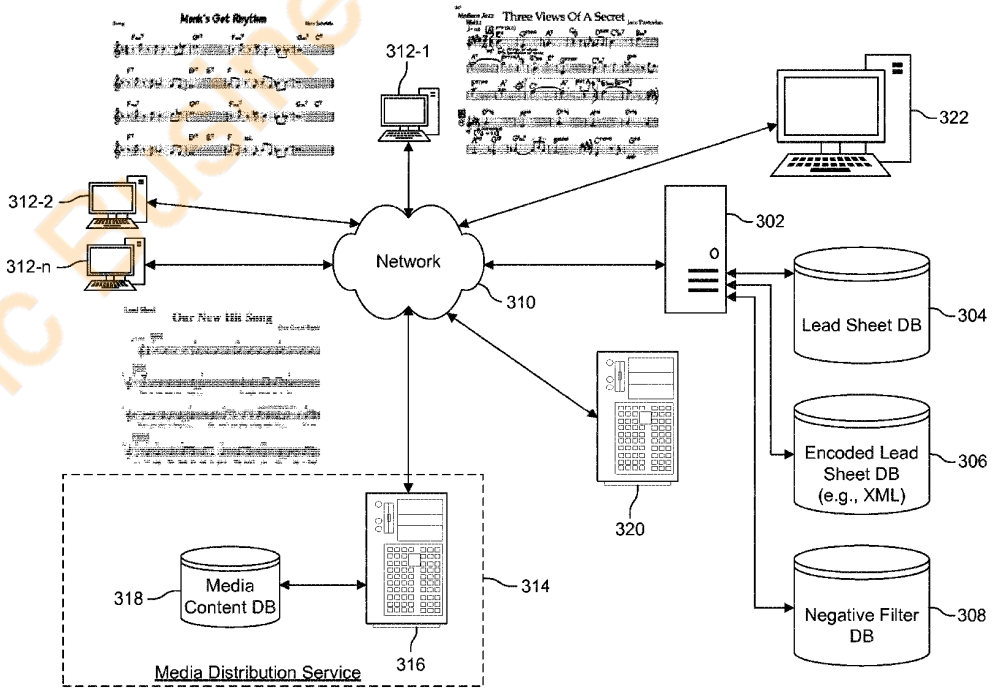


FIG. 3

EP 3 742 433 A1

Description

TECHNICAL FIELD

[0001] Example aspects described herein relate generally relate to plagiarism detection, and more particularly to a plagiarism risk detector and interface.

BACKGROUND

[0002] Plagiarism is the practice of taking the work or ideas of someone else and passing them off as one's own. It has been around practically as long as humans have produced works of art and research. One form of plagiarism, music plagiarism, is the use or close imitation of another author's music while representing it as one's own original work. Music plagiarism comes in various forms, generally summarized as sampling plagiarism, rhythm plagiarism and melody plagiarism.

[0003] Sampling plagiarism involves the re-use of recorded sounds or music excerpts in another song and can include manipulating the samples in, for example, pitch or tempo to fit the rhythm and tonality of a new song. Rhythm plagiarism is the general copying of the rhythm that is formed by a periodical pattern of accents in the amplitude envelopes of different frequency bands and can include a rhythm that has undergone a number of manipulations, such as time stretching, pitch shifting, re-sampling or even shuffling of individual beats. Melody plagiarism is the general copying of the melodic motive of a work, and can include a melodic motive that has been copied and then transposed to another key, slowed down, sped up or interpreted with different rhythmic accentuation.

[0004] When executed manually, plagiarism detection is usually performed by experts and lawyers. Manual detection of music plagiarism requires substantial effort, skill and excellent memory, and is generally known to be impractical. Software-assisted detection for text plagiarism on the other hand allows vast collections of documents to be compared to each other, making successful plagiarism detection much more likely.

[0005] Technology in the area of music plagiarism detection makes successful detection much more likely. Lee, J. et al., "Music Plagiarism Detection System", ITC-CSCC (2011), for example, describes a system that detects music plagiarism based on melodic similarity. Melody is obtained using the harmonic structure model, and similarity between two melodies is calculated using the edit distance. The system extracts melody from the input query and finds melodies in a database that are close to the query melody as potential melodies which the query has plagiarized. The *Lee et al.* system is implemented with a graphical user interface (GUI). The system is composed of four modules: (1) Melody Extraction Module, (2) Melody-to-MIDI Module, (3) Similarity Calculation Modules and 4) Common Subsequence Search Modules. The system receives as input a polyphonic music (PCM data) and outputs information of plagiarized music (music title, time, etc.).

[0006] Christian Dittmar, et al., "Audio Forensics Meets Music Information Retrieval - A Toolbox For Inspection Of Music Plagiarism" (2012) describes approaches to detecting and inspecting sampling plagiarism, rhythm plagiarism, and melody plagiarism. Sampling plagiarism inspection is detected by comparing a time-frequency representation of two music excerpts. A time-frequency representation of both music excerpts is compared by computing a magnitude spectrogram by means of STFT. Each spectral frame is then converted to a constant-Q representation by means of re-sampling to a logarithmically spaced frequency axis, yielding the spectrograms of original X_o and suspected plagiarism X_s respectively. A number of hypotheses f for the applied re-sampling factor is derived by computing the pair-wise ratio of the strongest periodicities in the energy envelope of X_o and X_s. In order to retrieve the occurrences of X_o inside X_s, it is re-sampled both in time and frequency according to each entry in $\boldsymbol{f}$ yielding X_o. Each X_o is shifted frame-wise along all frames of X_s and the accumulated, absolute difference $\boldsymbol{d}$ is computed between all corresponding time-frequency tiles.

[0007] Assuming only re-sampling and looping were applied, periodic minima will occur in $\boldsymbol{d}$. These correspond to the point, where an optimal matching can be found. At this point, it is also possible to subtract the energy of X_o from X_s, perform inverse short-time Fourier transform (STFT) and auralize the result. *Dittmar et al.* describe an alternative approach to detecting sampling plagiarism based on a decomposition of both X_o and X_s by means of Non-Negative Matrix Factorization (NMF).

[0008] *Dittmar et al.* also describe a rhythm plagiarism inspection technique that performs rhythmical source separation and tempo alignment. The rhythmical components of both X_o and X_s are again extracted by means of NMF. NMF is computed with large number of components that are, in turn, clustered. Features are extracted that indicate an assignment to a certain instrument. A measure is used for periodicity and all components that show a low percussiveness are removed. Afterwards, a clustering of the components performed. The assignment of components to each other is based on evaluating the correlation between the amplitude envelopes. A visualization can be presented in the plagiarism analyzer application for visual inspection by the user. The tempi of the sequences are aligned to each other. Finally, the extracted source from the original is compared to the extracted ones from the suspected plagiarism.

[0009] The score writing process typically involves writing music on a so-called "lead sheet". FIG. 1 illustrates an example prior art lead sheet. A lead sheet is a type of music score consisting of a monophonic melody 10 with associated chord labels 12, as shown in FIG. 1. Oftentimes, lead sheets also include lyrics 14 aligned with the melody.

[0010] A scorewriter, also sometimes referred to as a

music editor or music notation program, is software used with a computer for creating, editing and printing lead sheets. A scorewriter is to music notation what a word processor is to text, in that they both allow fast corrections (undo), flexible editing, easy sharing of electronic documents (via the Internet or compact storage media) and uniform layout.

[0011] While the above techniques for detecting plagiarism are significant improvements over manual approaches, they still require significant expertise and are not suited for operation by typical artists and composers, especially artists and composers who are interested in detecting plagiarism during the composition process. Moreover, the above-described plagiarism techniques are not integrated with the tools such artists and composers use to notate their works during the score writing process.

[0012] What is lacking from the prior art is a graphical user interface (GUI) that is more intuitive, more precise as to the portion of the work that may be considered plagiaristic, and that provides dynamic visual feedback in substantially real-time. Such a tool would allow artists to generate lead sheets more quickly and confidently by detecting and providing visual feedback as to whether any aspect of the work has a probability of being deemed plagiaristic. The GUI interface of *Lee et al.,* for example, provides an identification of the original song and other potentially similar songs. But the *Lee et al.* GUI does not provide specifics about what portion of the song might be the issue much less a GUI that visualizes the lead sheet in conjunction with plagiarism risk annotations.

[0013] *Dittmar et al.* provide more detailed visualizations of potential plagiaristic portions of a musical work. For example, the system of *Dittmar et al.* provides a visualization of the melody of an original work and the suspect plagiarism. However, similar to *Lee J. et al.*, the *Dittmar, C. et al.* system does not provide a visualization of the underlying lead sheet nor its particulars in a format that is more easily interpreted and navigated, for example, by artists, composers as well as publishers or right owners who want to protect their assets or otherwise need to assess to which extent a musical work infringes.

[0014] It would be useful to have a technology or service that provides risk assessment for a complete musical work in the form of a lead sheet, continuously and online during the composition process. It would be useful to be able to edit a lead sheet using a scorewriter while receiving fast and specific plagiarism detection information, for instance as annotations of the work in progress, stressing the degree of similarity of the composition with regards to a database of existing works.

SUMMARY

[0015] In an example embodiment, a method for testing a lead sheet for plagiarism is provided. The method includes receiving, at a plagiarism detector, a test lead sheet having a plurality of passages, the plagiarism detector having been trained on a plurality of preexisting encoded lead sheets; generating a set of annotations describing a level of plagiarism of a plurality of elements (e.g., chord sequence, subsequences, melodic fragments (i.e., notes), rhythm, harmony, etc.) of the test lead sheet in relation to the preexisting encoded lead sheets; and presenting (e.g., outputting) via an output device, the annotations.

[0016] In some embodiments the method further includes displaying the test lead sheet on the output device; and displaying the set of annotations on the output device by overlaying the set of annotations over the lead sheet. In an example embodiment, displaying the set of annotations can includes: overlaying each annotation of the set of annotations over any one of (i) a corresponding melodic fragment, (ii) a chord sequence, or (iii) a combination of (i) and (ii) depicted on the lead sheet.

[0017] Each annotation can indicate a portion of the plurality of elements and a level of plagiarism of the portion of the plurality of elements (e.g., "the chord sequence appears in many works of the database", "the melodic fragment appears to be completely new", "the melodic fragment appears in some works of the database").

[0018] In some embodiments, the method performs training of the plagiarism detector on a plurality of preexisting encoded lead sheets.

[0019] In yet other embodiments, the method performs: comparing each segment of the encoded test lead sheet to the plurality of preexisting encoded lead sheets; calculating a similarity value indicating the similarity of the segment of the encoded test lead sheet to a corresponding segment of the plurality of preexisting encoded lead sheets; and labeling as a match a segment of the encoded test lead sheet having a similarity value that meets a similarity threshold.

[0020] In some embodiments, the method performs storing at least one encoded filter element; comparing the at least one encoded filter element to the plurality of preexisting encoded lead sheets; and filtering out any segments of the plurality of preexisting encoded lead sheets that match.

[0021] In another example embodiment, a plagiarism detector for testing a lead sheet for plagiarism is provided. The plagiarism detector includes one or more processors configured to: receive an encoded test lead sheet representing a test lead sheet having a plurality of passages; generate a set of annotations describing a level of plagiarism of a plurality of elements of the encoded test lead sheet in relation to a plurality of preexisting encoded lead sheets; and cause an output device to present the annotations.

[0022] In some embodiments, the at least one processor can configured to: cause the output device to: display the test lead sheet; and display the set of annotations by overlaying the set of annotations over the lead sheet.

[0023] In yet other embodiments, the at least one processor is further configured to cause the output device to: overlay each annotation of the set of annotations over

any one of (i) a corresponding melodic fragment, (ii) a chord sequence, or (iii) a combination of (i) and (ii) depicted on the lead sheet.

[0024] In some embodiments, each annotation indicates a portion of the plurality of elements and a level of plagiarism of the portion of the plurality of elements.

[0025] In some embodiments, the at least one processor is further configured to: test the encoded test lead sheet against a model that has been trained on a plurality of preexisting encoded lead sheets.

[0026] In yet other embodiments, the at least one processor is further configured to: compare each segment of the encoded test lead sheet to the plurality of preexisting encoded lead sheets; calculate a similarity value indicating the similarity of the segment of the encoded test lead sheet to a corresponding segment of the plurality of preexisting encoded lead sheets; and label as a match a segment of the encoded test lead sheet having a similarity value that meets a similarity threshold.

[0027] Optionally, the plagiarism detector includes a negative filter database configured to store at least one encoded filter element. In this embodiment, the at least one processor further configured to: compare the at least one encoded filter element to the plurality of preexisting encoded lead sheets, and filter out any segments of the plurality of preexisting encoded lead sheets that match.

[0028] In yet another example embodiment, a non-transitory computer-readable medium having stored thereon one or more sequences of instructions for causing one or more processors to perform the methods described herein is provided.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] The features and advantages of the example embodiments of the invention presented herein will become more apparent from the detailed description set forth below when taken in conjunction with the following drawings.

FIG. 1 illustrates an example prior art lead sheet.

FIG. 2 illustrates an example score consisting of a single whole note and its representation in an electronic file format.

FIG. 3 illustrates a plagiarism risk detection system in accordance with an example embodiment of the present invention.

FIG. 4 depicts procedures for converting lead sheets to computer formatted lead sheet files and using the computer formatted lead sheet files to generate an output model in accordance with an example embodiment of the present invention.

FIG. 5A illustrates a procedure for testing a lead sheet to determine the probability that a component of the lead sheet plagiarizes an attributed work, in accordance with an example embodiment of the present invention.

FIG. 5B illustrates an example implementation of testing a lead sheet using a model in accordance with an example embodiment of the present invention.

FIG. 6 is a test lead sheet prepared using a scorewriter to be analyzed according to the example embodiments of the present invention.

FIG. 7 is an example of a test results overlay in accordance with an example embodiment of the present invention.

FIG. 8 illustrates an example screenshot of plagiarism-related information associated with the test lead sheet.

FIG. 9 is a block diagram for explaining additional details of a media control device with a single control input according to the example embodiments described herein.

DETAILED DESCRIPTION

[0030] The example embodiments of the invention presented herein are directed to methods, systems and computer program products for plagiarism risk assessment, which are now described herein in terms of an example cloud-based service for assessing the probability that a musical work in the form of a lead sheet is plagiaristic and presenting a graphical user interface identifying any potentially plagiaristic portions of the lead sheet along with relevant information. This description is not intended to limit the application of the example embodiments presented herein. In fact, after reading the following description, it will be apparent to one skilled in the relevant art(s) how to implement the following example embodiments in alternative embodiments (e.g., as a dedicated hardware device, and/or involving different types of music scores such as chord charts, and the like).

[0031] Generally, lead sheets are encoded in a computer format referred to herein as a music interchange format and the music interchange formatted lead sheets are uploaded to a database. The music interchange format thus contains one or more sequences of information representing the content of a lead sheet. A plagiarism risk assessment service (e.g., that operates a plagiarism risk detector) uses the uploaded music interchange formatted lead sheets for detecting possible plagiarism of a test lead sheet that has also been encoded in the music interchange format. The plagiarism risk assessment service returns a set of annotations describing which aspects of the test lead sheet are similar to existing lead sheets in the database.

[0032] In some embodiments, the plagiarism risk assessment service provides the annotations in real-time, and causes a graphical user interface (GUI) to display the annotations. The plagiarism risk assessment GUI can work in conjunction with a scorewriter application GUI. In some embodiments the plagiarism risk assessment GUI is combined with the scorewriter application GUI to provide annotations in substantially real time as the lead sheet is being composed. In some embodiments, the pla-

giarism risk assessment service is implemented in the form of a plugin of an existing scorewriter.

Electronically Formatting a Lead Sheet

[0033] Musical structure generally is the overall organization of a composition into sections, phrases, and patterns, very much like the organization of a text. Songs, for example, include sections, phrases and patterns that can often be further decomposed into elements that include melody, chord progression, rhythm, and lyrics.

[0034] Common Western music notation is a symbolic method of representing music for performers and listeners. Besides its use in publishing sheet music, musical scores and parts, the notation has been encoded in different computer formats, referred to herein as a music interchange formats. One example music interchange format is MusicXML which is an XML based format intended to be used with scorewriter tools to parse and manipulate a musical score. MusicXML is one type of music interchange format that is designed to allow the interchange of music notation data between and among music notation editing and publishing programs, as well as music scanning programs. While the example embodiments of the invention presented herein are described as using MusicXML it should be understood that other music interchange formats can be used instead of Music XML. Alternative embodiments can use different types of music interchange formats such as msf, RMTF, MIDI, abc, reativeMusicFile, FinaleFormat, ETF, RhapsodyFormat, EncoreFormat, Noteworthy, GuitarProFormat, TablEditFormat, SmartScore, and the like.

[0035] FIG. 2 illustrates an example prior art score 202 consisting of a single whole note and its representation in a music interchange format 204. In this example, the score 202 consists of a single whole note middle C in the key of C major on the Treble Clef and its representation using MusicXML code.

Plagiarism Risk Detection System

[0036] FIG. 3 illustrates a plagiarism risk detection system in accordance with an example embodiment of the present invention. A plagiarism risk detector 302 is coupled to one or more databases. In one example embodiment, plagiarism risk detector 302 is coupled to a lead sheet database 304. The lead sheet database 304 stores plural lead sheets in their native format. In another embodiment, plagiarism risk detector 302 is coupled to an encoded lead sheet database 306. An encoded lead sheet is a lead sheet that is encoded in a music interchange format. Encoded lead sheet database 306 stores encoded lead sheets (e.g., a corpus of lead sheets encoded in a music interchange format). In some embodiments, the plagiarism risk detector 302 includes at least one processor and a non-transitory memory storing instructions. When the instructions are executed by the at least one processor, the at least one processor performs the functions described herein.

[0037] In some embodiments, each encoded lead sheet is stored in encoded lead sheet database 306 as sequences S_1, S_2, ..., S_n, where n is an integer.

[0038] In an example implementation, fingerprinting is performed on the segments of the sequences using a fingerprinting algorithm. Generally, a fingerprinting algorithm maps the data contained in the sequences (e.g., segments of the sequences) to, for example, shorter text strings. Such shorter text strings are known as fingerprints. These fingerprints are unique identifiers for their corresponding data and/or files. Now known or future developed mechanisms for fingerprinting and matching encoded test lead sheets to a corpus of encoded lead sheets stored in encoded lead sheet database 306 can be used.

[0039] In yet another example embodiment, plagiarism risk detector 302 is coupled to a negative filter database 308. In some embodiment, such elements are also encoded in a music exchange format and are referred to herein as encoded filter elements. Negative filter database 308 stores elements of musical scores that are viewed as non-plagiaristic. Negative filter database 308 is used, for example, to filter out matches that are permissible uses, common features of musical scores, or other sections, phrases, and/or patterns (e.g., melodies, chord progressions, rhythms, and lyrics) that are common or otherwise would report false positives for plagiarism. In an example implementation, a negative filter database 308 stores encoded filter elements F_1, F_2, ..., F_x, where x is an integer. The filtering process involves comparing segments of a collection of source sequences S_1, S_2, ..., S_n, where n is an integer (e.g., representing encoded lead sheets stored in an encoded lead sheet database 306) with segments of sequences of encoded filter elements Fi, F_2, ..., F_x, where x is an integer. The matched segments (e.g., the segments that are similar or substantially similar) are, in turn, filtered out. That is, the matched segments are filtered and not compared to a test lead sheet.

[0040] In an example embodiment, fingerprinting is performed on segments of sequences of the encoded filter elements stored in negative filter database 308. Fingerprinting is also performed on the segments of source sequences stored in encoded lead sheet database 306. In this embodiment, one or more fingerprints of the encoded filter elements are compared against the fingerprints of the encoded lead sheets. This reduces the amount of processing resources that need to be used to test an encoded test lead sheet by reducing the test data set that the encoded test lead sheet is compared against.

[0041] As shown in FIG. 3, plagiarism risk detector 302 is coupled to various sources of lead sheets 312-1, 312-2, ..., 312-n via a network 310. In addition or alternatively, plagiarism risk detector 302 can be coupled to a media distribution service 314 that includes a music distribution server 316 and a media content database 318 that stores media content items. The media distribu-

tion service 314 can provide streams of media content or media content items for downloading to plagiarism risk detector 302. In one embodiment plagiarism risk detector 302 converts the music content of the media content items into encoded lead sheets. In turn, the encoded lead sheets are stored in encoded lead sheet database 306 for later processing.

[0042] In another example embodiment, a notation service 320 converts media content (e.g., songs) from, for example media distribution service 314 into encoded lead sheets and supplies the encoded lead sheets to encoded lead sheet database 306 for later processing.

[0043] As explained above segments of a collection of source sequences S1, S2, ..., S_n, where ***n*** is an integer, representing encoded lead sheets are stored in the encoded lead sheet database 306. In some embodiments, fingerprints of the segments can be stored, for example to decrease the amount of time it takes to compare the segments, to increase the ability to make accurate comparisons, and to reduce processing resources.

[0044] Plagiarism risk detector 302 uses the encoded lead sheets stored in encoded lead sheet database 306 to detect possible plagiarism and provide a set of annotations describing which elements of a test lead sheets are similar to existing lead sheets in the encoded lead sheet database 306.

[0045] In some embodiments, plagiarism risk detector 302 is communicatively coupled to client device 322. In one embodiment, Plagiarism risk detector 302 is coupled to client device 322 via network 310. Client device 322 includes one or more processors and a non-transitory memory device storing an integrated scorewriting and plagiarism detection application, which when executed by the one or more processors causes the client device to operate as an integrated scorewriter and plagiarism detector.

Lead Sheet Conversion and Output Model Generation Procedures

[0046] FIG. 4 depicts a procedure for converting lead sheets to computer formatted lead sheet files 420 and a procedure for using the computer formatted lead sheet files to generate an output model 430 in accordance with an example embodiments of the present invention. At block S422, lead sheet encoding procedure 420 receives lead sheets in their native format and in block S424 encodes the lead sheets to generate a computer formatted lead sheet files, referred to herein as encoded lead sheets. In turn, the encoded lead sheets are stored in an encoded lead sheet database (e.g., FIG. 1, 306), as shown in block S426.

[0047] As described above, in some embodiments, the computer format used to generate computer formatted lead sheet files is a music interchange format. In some example embodiments lead sheet encoding procedure 420 transmits the encoded lead sheets to another service or system for further processing. Lead sheet learning procedure 430 is such a processing service.

[0048] Lead sheet learning procedure 430 retrieves the encoded lead sheet files as shown in block S432, performs a learning algorithm on the computer formatted lead sheet files S434, and generates an output model S436. The machine learning algorithm that is used to generate the output model is not limited to any machine algorithm implementation. Indeed, in some embodiments, combining multiple base learners can result in improved prediction performance. Those skilled in the art will appreciate that now known or future developed learning algorithms can be used to train the output model.

Lead Sheet Plagiarism Detection Procedure

[0049] FIG. 5A illustrates a procedure 450 for testing a lead sheet to determine the probability that a component of the lead sheet plagiarizes an attributed work, in accordance with an example embodiment of the present invention.

[0050] In block S452, an encoded test lead sheet is received. The encoded test lead sheet 502 is also sometimes referred to as a query lead sheet.

[0051] If a lead sheet to be tested is not already in a music interchange formats, the lead sheet is converted into an encoded lead sheet file 502.

[0052] In the example embodiment depicted in FIG. 5A, the encoded test lead sheet 502 is in a music interchange format as described above.

[0053] An example test lead sheet is illustrated in FIG. 6. In the example shown in FIG. 6, the test lead sheet is a lead sheet that is prepared using a scorewriter application. The scorewriter application saves an encoded version of the test lead sheet (i.e., the encoded test lead sheet) in a memory store (either locally, e.g., on a disk drive or remotely, e.g., on the cloud). In some embodiments, the encoded test lead sheet is updated in real time as changes to the lead sheet are being made through the use of the socrewriter application.

[0054] In block S454, the test lead sheet is evaluated against a corpus of encoded lead sheets. This can be accomplished in a number of ways.

[0055] FIG. 5B illustrates an example implementation of testing a lead sheet using a model (block S454 of FIG. 5A) in accordance with an example embodiment of the present invention.

[0056] In some embodiments, the encoded test lead sheet is formatted as a sequence (e.g., a digitized chord sequence, a digitized subsequence, and the like). Referring to FIG. 5B, such an encoded test lead sheet is also referred to as a target sequence **T** 502. Given the target sequence **T** 502 and given a collection of source sequences S_1, S_2, ..., S_n (e.g., representing preexisting encoded lead sheets stored in an encoded lead sheet database 306), a search is performed for every segment **Seg of T** 502 for a list of all segments of all sequences **Si**(i=1,2,.., n, where ***i*** is an integer) that are similar to **Seg**, using, for example, a similarity measure, as shown

in block S545-1 of FIG. 5B. A similarity measurement, e.g., performed by a processor referred to for convenience as a similarity test processor, generates a quantity that reflects the strength of a relationship between two objects or two features, referred to herein as a similarity value. Here the similarity measurement generates a quantity that reflects the strength of the relationship between a segment (Seg) of an encoded test lead sheet to one or more segments of preexisting encoded lead sheets stored in encoded lead sheet database 306. This similarity measurement can be computed in many different ways. For example, the similarity measurement can be computed by performing a sequence alignment algorithm such as the Smith-Waterman algorithm, as shown in block S545-2.

[0057] In some embodiments, a method performs calculating a similarity value indicating the similarity of the segment of the encoded test lead sheet to a corresponding segment of the plurality of preexisting encoded lead sheets and identifying a segment of the encoded test lead sheet having a similarity value that meets a similarity threshold. The segment of the encoded test lead sheet having a similarity value that meets the similarity threshold is labeled as a match (i.e., as potentially plagiaristic), as shown in block S545-3.

[0058] With this information, the segments of the target sequence which have the highest number of matches M (M, where M is an integer) in the source collection can be identified as being potentially plagiaristic.

[0059] In some embodiments, the music score being composed, e.g., the target sequence T can be rendered as an audio file (e.g. using a MIDI synthesizer). Then sampling detection methods can be used to detect similar audio segments in the source collection (themselves rendered as audio files).

[0060] As used herein a musical element refers to sections, phrases, and patterns. With respect to songs, for example, the term musical element includes sections, phrases and patterns that can be further decomposed into elements that include melody, chord progression, rhythm, and lyrics.

[0061] Referring back to FIG. 5A, in block S456 test results are generated. In block S458, a user interface graphical overlay is generated based on the test results and overlaid onto the test lead sheet. FIG. 7 is an example of a test results overlay in accordance with an example embodiment of the present invention. I this example, the annotations illustrate whether chord sequences of the test lead sheet match the preexisting works. In one example implementation, a high probability of plagiarism message 702 is presented to the operator. In this example, the message states that a particular chord sequence in measure 3-5 of the test lead sheet appear in many works. In some embodiments, a message can be presented to the operator indicating that there appears to be no match. For example, as shown in FIG. 7, the interface is configured to present a message stating that a particular melodic fragment does not appear to be found 704. Yet another message can indicate to the operator that only some matches were found (e.g., less than a predetermined threshold). In the example depicted in FIG. 7, a melodic fragment at measures 7-9 of the test lead sheet has been flagged as being matched to some works 706.

[0062] In turn, in block S460, the test result user interface (UI) overlay is displayed to appear on top of (e.g., overlaid over) the lead sheet notation. At block S462 a determination is made whether a test result user interface overlay has been selected. If so, then at block S464 additional information is rendered onto the display.

[0063] FIG. 8 illustrates an example screenshot 800 of plagiarism-related information associated with the test lead sheet, in accordance with an embodiment of the present invention. As shown in FIG. 8, the music annotations that are potentially plagiaristic are identified in terms of their locations 510. This can be, for example, the measures in the music score as depicted on the test lead sheet being generated using the scorewriter application. In some embodiments, the particular measures 510 that are potentially plagiaristic corresponds to the segments of sequences of the encoded test lead sheet that matched the encoded lead sheets that are stored in encoded lead sheet database 306. Also depicted in the test results summary is the type of plagiarism 520 that was detected (e.g., sampling, melody, rhythm, chord sequence).

[0064] In some embodiments a link to the media content item that might be infringed (e.g., a track of an album) is provided so that an operator can quickly select the link to listen to the potentially plagiarized work. The links (or the track identifiers) are illustrated here by track identifier 530. However, other forms of identification can be used (E.g., name of song). The number of works 540 potentially plagiarized can also be presented via interface 800.

[0065] It will be recognized by those skilled in the art that additional information can be provided via the user interface. For example, a plagiarism probability value (not shown) of the potential plagiarism can be displayed. The calculation can be based on the similarity value. Those skilled in the art will recognize that additional information can be displayed and still be within the scope of the invention.

[0066] The example embodiments presented herein may be provided as a computer program product, or software, that may include an article of manufacture on a machine accessible or machine readable medium having instructions. The instructions on the non-transitory machine accessible machine readable or computer-readable medium may be used to program a computer system or other electronic device. The machine or computer-readable medium may include, but is not limited to, optical disks, CD-ROMs, and magneto-optical disks or other type of media/machine-readable medium suitable for storing or transmitting electronic instructions. The techniques described herein are not limited to any particular software configuration. They may find applicability in any

computing or processing environment. The terms "computer-readable", "machine accessible medium" or "machine readable medium" used herein shall include any medium that is capable of storing, encoding, or transmitting a sequence of instructions for execution by the machine and that cause the machine to perform any one of the methods described herein. Furthermore, it is common in the art to speak of software, in one form or another (e.g., program, procedure, process, application, module, unit, logic, and so on) as taking an action or causing a result. Such expressions are merely a shorthand way of stating that the execution of the software by a processing system causes the processor to perform an action to produce a result.

[0067] Portions of the example embodiments of the invention may be conveniently implemented by using a conventional general purpose computer, a specialized digital computer and/or a microprocessor programmed according to the teachings of the present disclosure, as is apparent to those skilled in the computer art. Appropriate software coding may readily be prepared by skilled programmers based on the teachings of the present disclosure.

[0068] Some embodiments may also be implemented by the preparation of application-specific integrated circuits, field programmable gate arrays, or by interconnecting an appropriate network of conventional component circuits.

[0069] Some embodiments include a computer program product. The computer program product may be a storage medium or media having instructions stored thereon or therein which can be used to control, or cause, a computer to perform any of the procedures of the example embodiments of the invention. The storage medium may include without limitation an optical disc, a Blu-ray Disc, a DVD, a CD or CD-ROM, a micro-drive, a magneto-optical disk, a ROM, a RAM, an EPROM, an EEPROM, a DRAM, a VRAM, a flash memory, a flash card, a magnetic card, an optical card, nanosystems, a molecular memory integrated circuit, a RAID, remote data storage/archive/warehousing, and/or any other type of device suitable for storing instructions and/or data.

[0070] Stored on any one of the computer readable medium or media, some implementations include software for controlling both the hardware of the general and/or special computer or microprocessor, and for enabling the computer or microprocessor to interact with a human user or other mechanism utilizing the results of the example embodiments of the invention. Such software may include without limitation device drivers, operating systems, and user applications. Ultimately, such computer readable media further includes software for performing example aspects of the invention, as described above.

[0071] Included in the programming and/or software of the general and/or special purpose computer or microprocessor are software modules for implementing the procedures described above.

[0072] FIG. 9 is a block diagram for explaining further details of a plagiarism risk detector 302 in accordance with some of the example embodiments described herein. Plagiarism risk detector 302 includes a processor device 910, a main memory 925, and an interconnect bus 905. The processor device 910 may include without limitation a single microprocessor, or may include a plurality of microprocessors for configuring the plagiarism risk detector 302 as a multiprocessor system. The main memory 925 stores, among other things, instructions and/or data for execution by the processor device 910. The main memory 925 may include banks of dynamic random access memory (DRAM), as well as cache memory.

[0073] The plagiarism risk detector 302 may further include a mass storage device 930, peripheral device(s) 940, portable non-transitory storage medium device(s) 950, input control device(s) 980, a graphics subsystem 960, and/or an output display interface 970. For explanatory purposes, all components in the plagiarism risk detector 302 are shown in FIG. 9 as being coupled via the bus 905. However, the plagiarism risk detector 302 is not so limited. Elements of the plagiarism risk detector 302 may be coupled via one or more data transport means. For example, the processor device 910 and/or the main memory 925 may be coupled via a local microprocessor bus. The mass storage device 930, peripheral device(s) 940, portable storage medium device(s) 950, and/or graphics subsystem 960 may be coupled via one or more input/output (I/O) buses. The mass storage device 930 may be a nonvolatile storage device for storing data and/or instructions for use by the processor device 910. The mass storage device 930 may be implemented, for example, with a magnetic disk drive or an optical disk drive. In a software embodiment, the mass storage device 930 is configured for loading contents of the mass storage device 930 into the main memory 925.

[0074] Mass storage device 930 additionally stores code for executing the similarity measurement (e.g., similarity test processor) 931, test result generator 932, test results overlay generator 933, test result user interface UI 934, and negative filter 935. Similarity test processor 931 receives encoded lead sheets in a and performs a similarity measurement to determine whether any segments of sequences of the test lead sheet potentially plagiarizes any segments of sequences of preexisting encoded lead sheets. Test result generator 932 generates the test results based on a comparison of the test lead sheet against the corpus of test lead sheets. Test result user interface (UI) overlay generator 933 performs the rendering of the test results user interface overlay onto a screen, and Test results UI receives input and output from a client device on which a test music score is generated. Negative filter 935 performs negative filtering to filter out matches that are permissible uses, common features of musical scores, or other sections, phrases, and/or patterns (e.g., melodies, chord progressions, rhythms, and lyrics) that are common or otherwise would report false positives for plagiarism.

[0075] The portable storage medium device 950 operates in conjunction with a nonvolatile portable storage medium, such as, for example, flash memory, to input and output data and code to and from the plagiarism risk detector 302. In some embodiments, the software for storing information may be stored on a portable storage medium, and may be inputted into the plagiarism risk detector 302 via the portable storage medium device 950. The peripheral device(s) 940 may include any type of computer support device, such as, for example, an input/output (I/O) interface configured to add additional functionality to the plagiarism detector 302. For example, the peripheral device(s) 940 may include a network interface card for interfacing the plagiarism risk detector 302 with a network 920.

[0076] The input control device(s) 980 provide a portion of the user interface for a user of the plagiarism risk detector 302. The input control device(s) 980 may include a keypad and/or a cursor control device. The keypad may be configured for inputting alphanumeric characters and/or other key information. The cursor control device may include, for example, a handheld controller or mouse, a trackball, a stylus, and/or cursor direction keys. The plagiarism risk detector 302 may include an optional graphics subsystem 960 and output display 970 to display textual and graphical information. The output display 970 may include a display such as a CSTN (Color Super Twisted Nematic), TFT (Thin Film Transistor), TFD (Thin Film Diode), OLED (Organic Light-Emitting Diode), AMOLED display (Activematrix organic light-emitting diode), and/or liquid crystal display (LCD)-type displays. The displays can also be touchscreen displays, such as capacitive and resistive-type touchscreen displays.

[0077] The graphics subsystem 960 receives textual and graphical information, and processes the information for output to the output display 970. Input control devices 980 can control the operation and various functions of the plagiarism risk detector 302.

[0078] Input control devices 980 can include any components, circuitry, or logic operative to drive the functionality of the plagiarism detector 302. For example, input control device(s) 980 can include one or more processors acting under the control of an application.

[0079] Also shown FIG. 9 is media playback device 990. As described above, the plagiarism risk detector 302 can have its own media playback component or functionality or a media playback device 990 can be integrated into the plagiarism risk detector 302.

[0080] Various operations and processes described herein can be performed by the cooperation of two or more devices, systems, processes, or combinations thereof.

[0081] While various example embodiments of the invention have been described above, it should be understood that they have been presented by way of example, and not limitation. It is apparent to persons skilled in the relevant art(s) that various changes in form and detail can be made therein. Thus, the disclosure should not be limited by any of the above described example embodiments, but should be defined only in accordance with the following claims and their equivalents. Further, the Abstract is not intended to be limiting as to the scope of the example embodiments presented herein in any way. It is also to be understood that the procedures recited in the claims need not be performed in the order presented.

Claims

1. A method for testing a lead sheet for plagiarism, comprising the steps of:

   receiving, at a plagiarism detector, an encoded test lead sheet representing a test lead sheet having a plurality of passages;
   generating a set of annotations describing a level of plagiarism of a plurality of elements of the encoded test lead sheet in relation to a plurality of preexisting encoded lead sheets; and
   presenting via an output device, the annotations.

2. The method according to Claim 1, further comprising the steps of:

   displaying the test lead sheet on the output device; and
   displaying the set of annotations on the output device by overlaying the set of annotations over the lead sheet.

3. The method according to Claim 2, wherein displaying the set of annotations includes:
   overlaying each annotation of the set of annotations over any one of (i) a corresponding melodic fragment, (ii) a chord sequence, or (iii) a combination of (i) and (ii) depicted on the lead sheet.

4. The method according to any preceding claim, wherein each annotation indicates a portion of the plurality of elements and a level of plagiarism of the portion of the plurality of elements.

5. The method according to any preceding claim, further comprising the step of
   training the plagiarism detector on the plurality of preexisting encoded lead sheets.

6. The method according to any preceding claim, further comprising the steps of:

   comparing each segment of the encoded test lead sheet to the plurality of preexisting encoded lead sheets;
   calculating a similarity value indicating the similarity of the segment of the encoded test lead

sheet to a corresponding segment of the plurality of preexisting encoded lead sheets; and
labeling as a match a segment of the encoded test lead sheet having a similarity value that meets a similarity threshold.

7. The method according to any preceding claim, further comprising the steps of:

storing at least one encoded filter element;
comparing the at least one encoded filter element to the plurality of preexisting encoded lead sheets; and
filtering out any segments of the plurality of preexisting encoded lead sheets that match.

8. A plagiarism detector for testing a lead sheet for plagiarism, comprising:
at least one processor operable to:

receive an encoded test lead sheet representing a test lead sheet having a plurality of passages;
generate a set of annotations describing a level of plagiarism of a plurality of elements of the encoded test lead sheet in relation to a plurality of preexisting encoded lead sheets; and
cause an output device to present the annotations.

9. The plagiarism detector according to Claim 8, the at least one processor further configured to:
cause the output device to:

display the test lead sheet; and
display the set of annotations by overlaying the set of annotations over the lead sheet.

10. The plagiarism detector according to Claim 9, the at least one processor further configured to cause the output device to:
overlay each annotation of the set of annotations over any one of (i) a corresponding melodic fragment, (ii) a chord sequence, or (iii) a combination of (i) and (ii) depicted on the lead sheet.

11. The plagiarism detector according to any claim 8-10, wherein each annotation indicates a portion of the plurality of elements and a level of plagiarism of the portion of the plurality of elements.

12. The plagiarism detector according to any claim 8-11, the at least one processor further configured to:
test the encoded test lead sheet against a model that has been trained on the plurality of preexisting encoded lead sheets.

13. The plagiarism detector according to any claim 8-12, the at least one processor further configured to:

compare each segment of the encoded test lead sheet to the plurality of preexisting encoded lead sheets;
calculate a similarity value indicating the similarity of the segment of the encoded test lead sheet to a corresponding segment of the plurality of preexisting encoded lead sheets; and
label as a match a segment of the encoded test lead sheet having a similarity value that meets a similarity threshold.

14. The plagiarism detector according to any claim 8-13, further comprising:

a negative filter database configured to store at least one encoded filter element; and
the at least one processor further configured to:

compare the at least one encoded filter element to the plurality of preexisting encoded lead sheets, and
filter out any segments of the plurality of preexisting encoded lead sheets that match.

15. A non-transitory computer-readable medium having stored thereon one or more sequences of instructions for causing one or more processors to perform:

receiving, at a plagiarism detector, an encoded test lead sheet representing a test lead sheet having a plurality of passages;
generating a set of annotations describing a level of plagiarism of a plurality of elements of the encoded test lead sheet in relation to a plurality of preexisting encoded lead sheets; and
presenting via an output device, the annotations.

Amended claims in accordance with Rule 137(2) EPC.

1. A method for testing a lead sheet for plagiarism, comprising the steps of:

receiving, at a plagiarism detector, an encoded test lead sheet representing a test lead sheet having a plurality of passages;
generating a set of annotations describing a level of plagiarism of a plurality of elements of the encoded test lead sheet in relation to a plurality of preexisting encoded lead sheets, comprising computing a similarity measurement by performing a Smith-Waterman algorithm; and
presenting via an output device, the annotations.

2. The method according to Claim 1, further comprising

the steps of:

displaying the test lead sheet on the output device; and
displaying the set of annotations on the output device by overlaying the set of annotations over the lead sheet.

3. The method according to Claim 2, wherein displaying the set of annotations includes:
overlaying each annotation of the set of annotations over any one of (i) a corresponding melodic fragment, (ii) a chord sequence, or (iii) a combination of (i) and (ii) depicted on the lead sheet.

4. The method according to any preceding claim, wherein each annotation indicates a portion of the plurality of elements and a level of plagiarism of the portion of the plurality of elements.

5. The method according to any preceding claim, further comprising the step of training the plagiarism detector on the plurality of preexisting encoded lead sheets.

6. The method according to any preceding claim, further comprising the steps of:

comparing each segment of the encoded test lead sheet to the plurality of preexisting encoded lead sheets;
calculating a similarity value indicating the similarity of the segment of the encoded test lead sheet to a corresponding segment of the plurality of preexisting encoded lead sheets; and
labeling as a match a segment of the encoded test lead sheet having a similarity value that meets a similarity threshold.

7. The method according to any preceding claim, further comprising the steps of:

storing at least one encoded filter element;
comparing the at least one encoded filter element to the plurality of preexisting encoded lead sheets; and
filtering out any segments of the plurality of preexisting encoded lead sheets that match.

8. A plagiarism detector for testing a lead sheet for plagiarism, comprising:
at least one processor operable to:

receive an encoded test lead sheet representing a test lead sheet having a plurality of passages;
generate a set of annotations describing a level of plagiarism of a plurality of elements of the encoded test lead sheet in relation to a plurality of preexisting encoded lead sheets, comprising computing a similarity measurement by performing a Smith-Waterman algorithm; and
cause an output device to present the annotations.

9. The plagiarism detector according to Claim 8, the at least one processor further configured to:
cause the output device to:

display the test lead sheet; and
display the set of annotations by overlaying the set of annotations over the lead sheet.

10. The plagiarism detector according to Claim 9, the at least one processor further configured to cause the output device to:
overlay each annotation of the set of annotations over any one of (i) a corresponding melodic fragment, (ii) a chord sequence, or (iii) a combination of (i) and (ii) depicted on the lead sheet.

11. The plagiarism detector according to any claim 8-10, wherein each annotation indicates a portion of the plurality of elements and a level of plagiarism of the portion of the plurality of elements.

12. The plagiarism detector according to any claim 8-11, the at least one processor further configured to:
test the encoded test lead sheet against a model that has been trained on the plurality of preexisting encoded lead sheets.

13. The plagiarism detector according to any claim 8-12, the at least one processor further configured to:

compare each segment of the encoded test lead sheet to the plurality of preexisting encoded lead sheets;
calculate a similarity value indicating the similarity of the segment of the encoded test lead sheet to a corresponding segment of the plurality of preexisting encoded lead sheets; and
label as a match a segment of the encoded test lead sheet having a similarity value that meets a similarity threshold.

14. The plagiarism detector according to any claim 8-13, further comprising:

a negative filter database configured to store at least one encoded filter element; and
the at least one processor further configured to:

compare the at least one encoded filter element to the plurality of preexisting encoded lead sheets, and
filter out any segments of the plurality of pre-

existing encoded lead sheets that match.

15. A non-transitory computer-readable medium having stored thereon one or more sequences of instructions for causing one or more processors to perform:

receiving, at a plagiarism detector, an encoded test lead sheet representing a test lead sheet having a plurality of passages;
generating a set of annotations describing a level of plagiarism of a plurality of elements of the encoded test lead sheet in relation to a plurality of preexisting encoded lead sheets; and
presenting via an output device, the annotations.

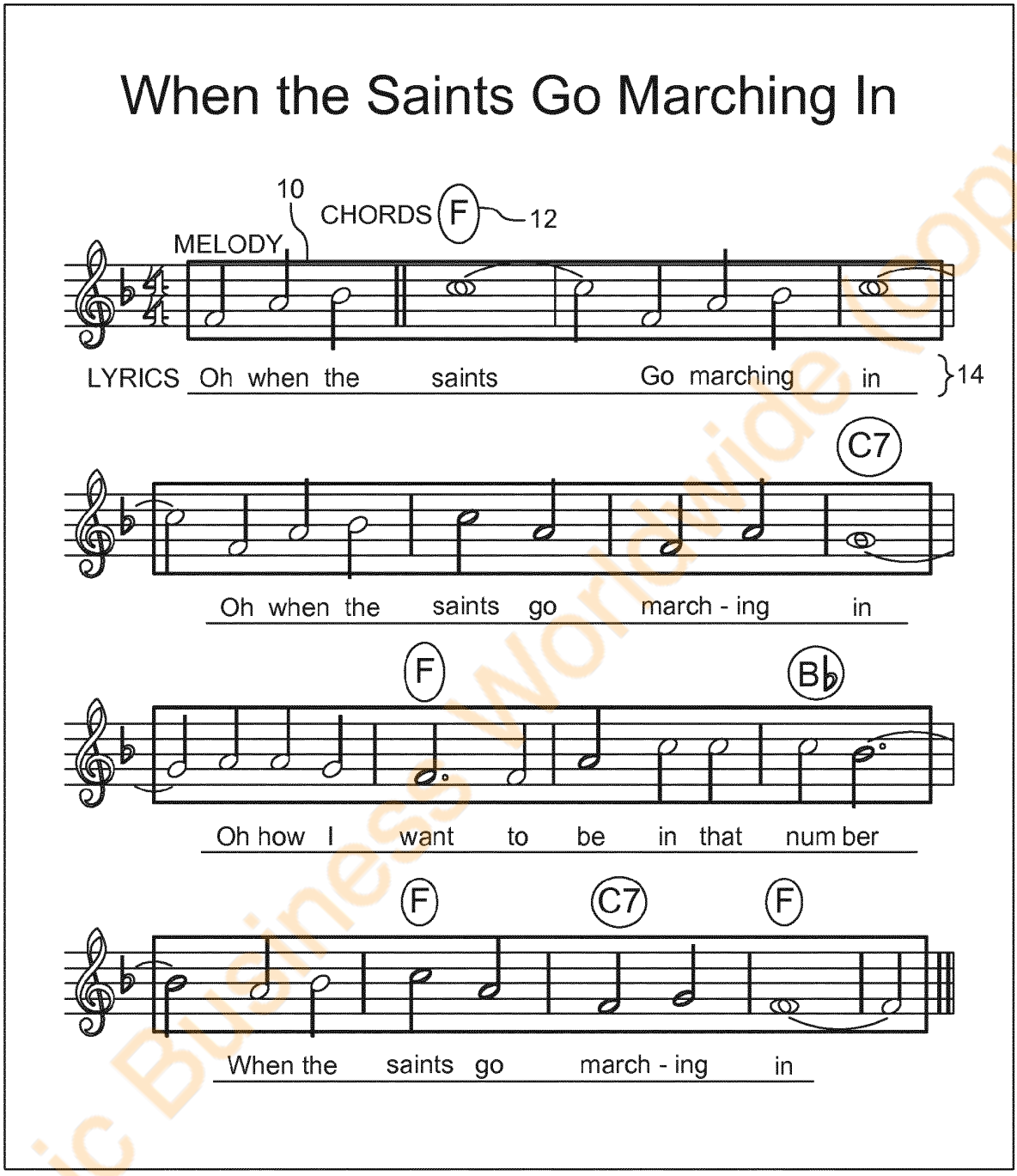
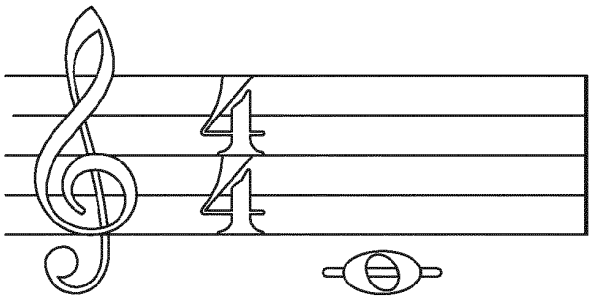


FIG. 1
PRIOR ART



202

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE score-partwise PUBLIC
        "-//Recordare//DTD MusicXML 3.1 Partwise//EN"
        "http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise version="3.1">
   <part-list>
      <score-part id="P1">
         <part-name>Music</part-name>
      </score-part>
   </part-list>
   <part id="P1">
      <measure number="1">
         <attributes>
            <divisions>1</divisions>
            <key>
               <fifths>0</fifths>
            </key>
            <time>
               <beats>4</beats>
               <beat-type>4</beat-type>
            </time>
            <clef>
               <sign>G</sign>
               <line>2</line>
            </clef>
         </attributes>
         <note>
            <pitch>
               <step>C</step>
               <octave>4</octave>
            </pitch>
            <duration>4</duration>
            <type>whole</type>
         </note>
      </measure>
   </part>
</score-partwise>
```

204

FIG. 2
PRIOR ART

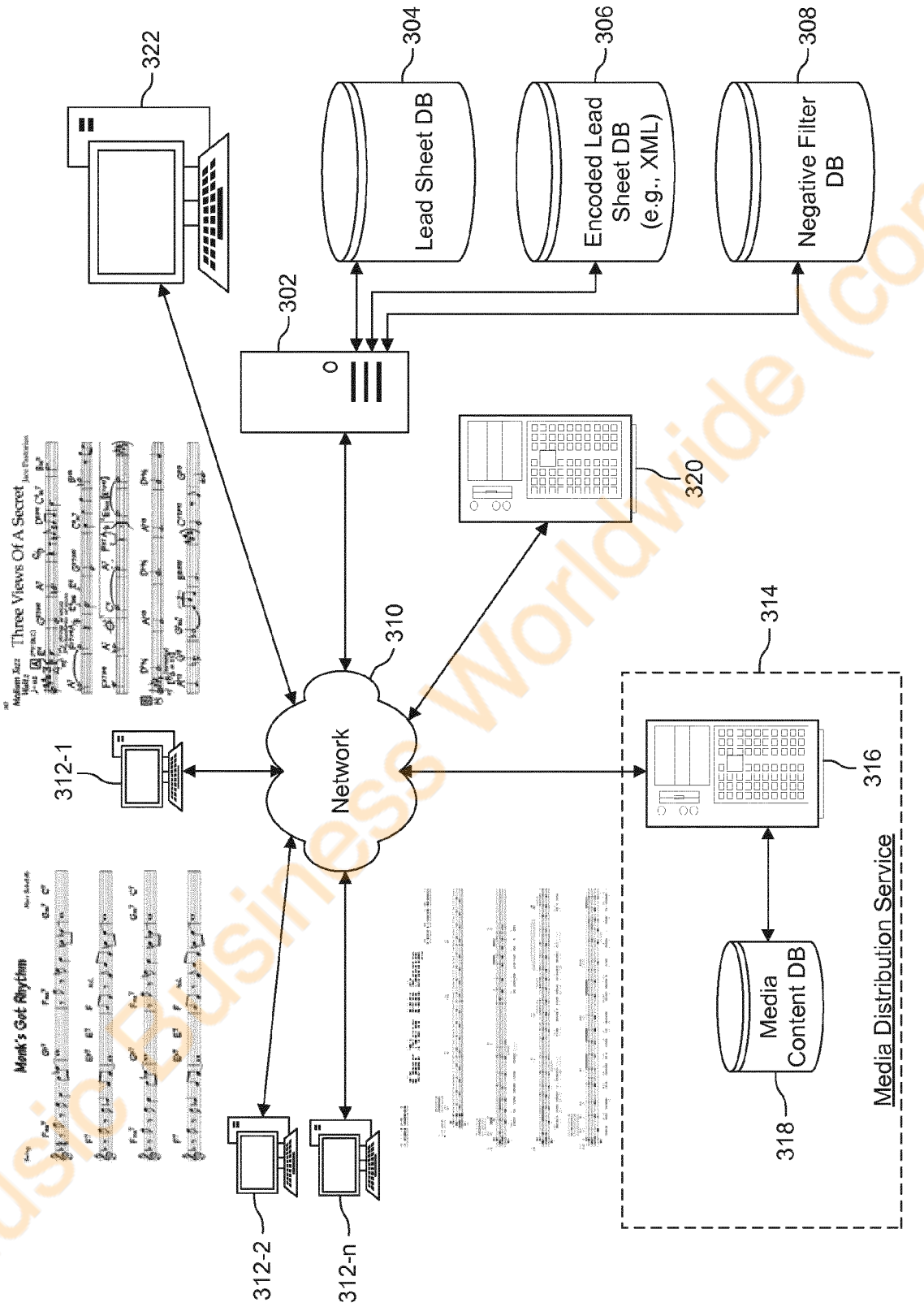


FIG. 3

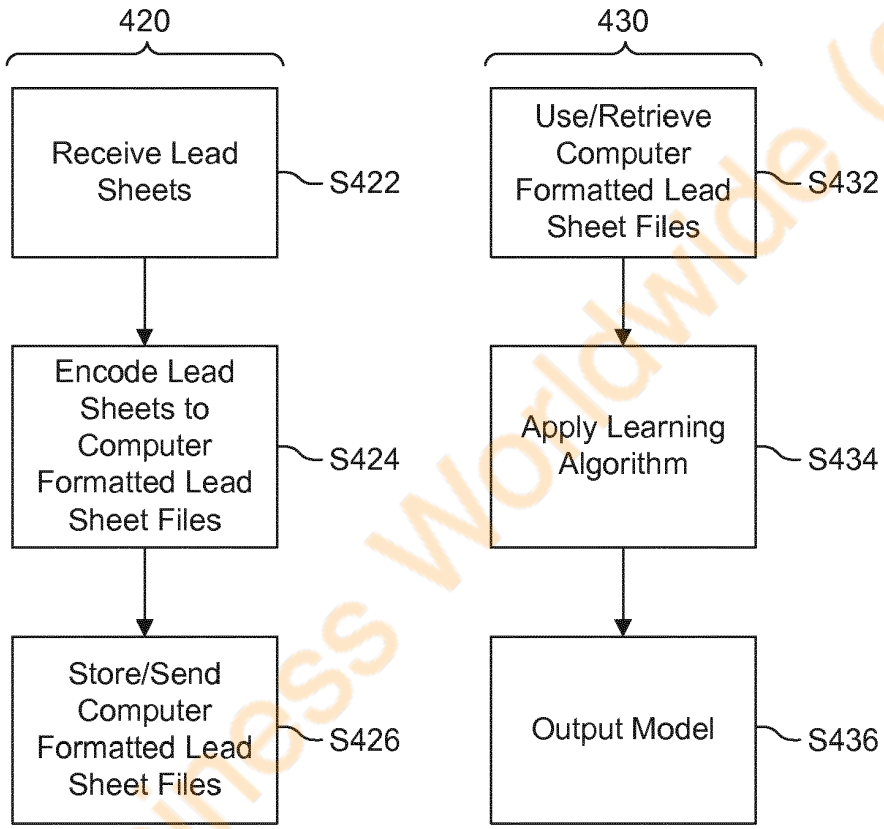


FIG. 4

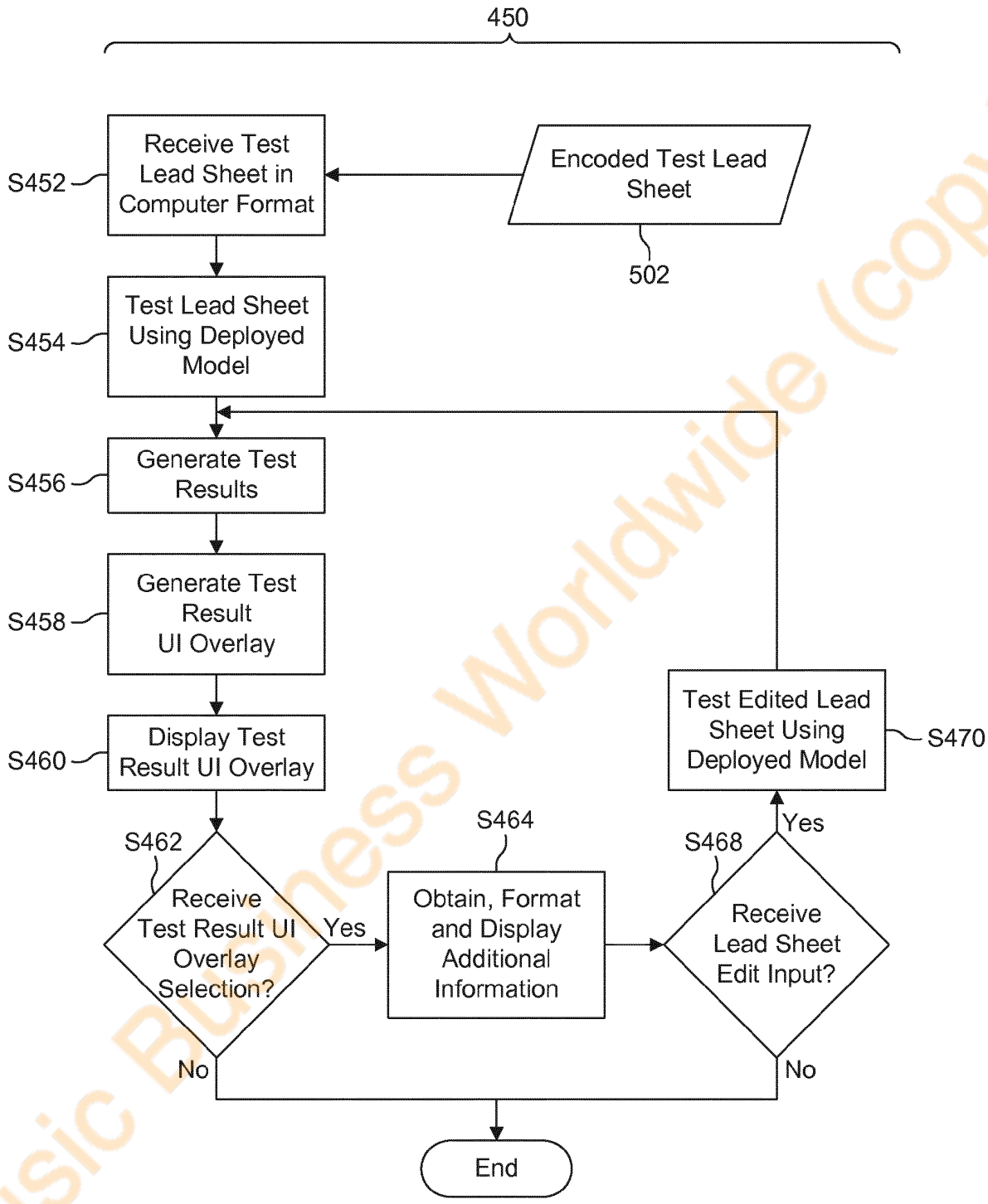


FIG. 5A

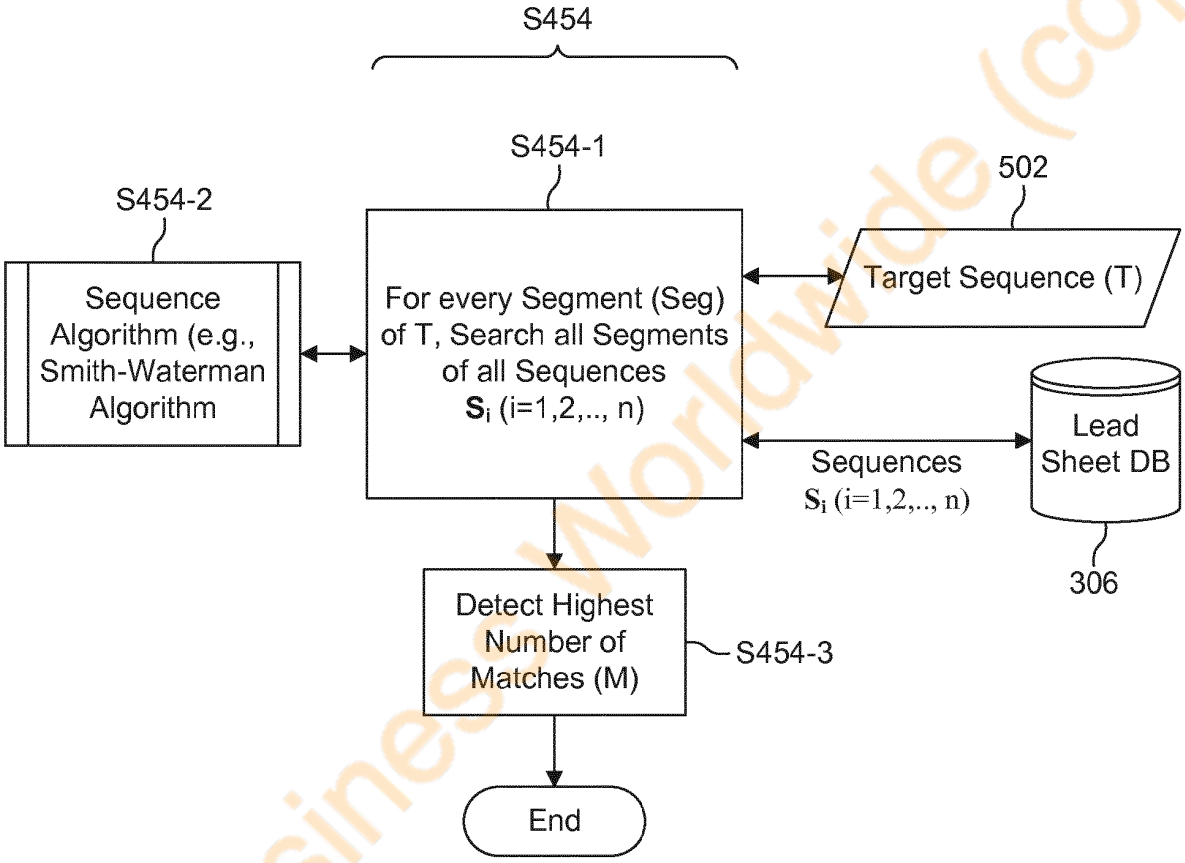


FIG. 5B

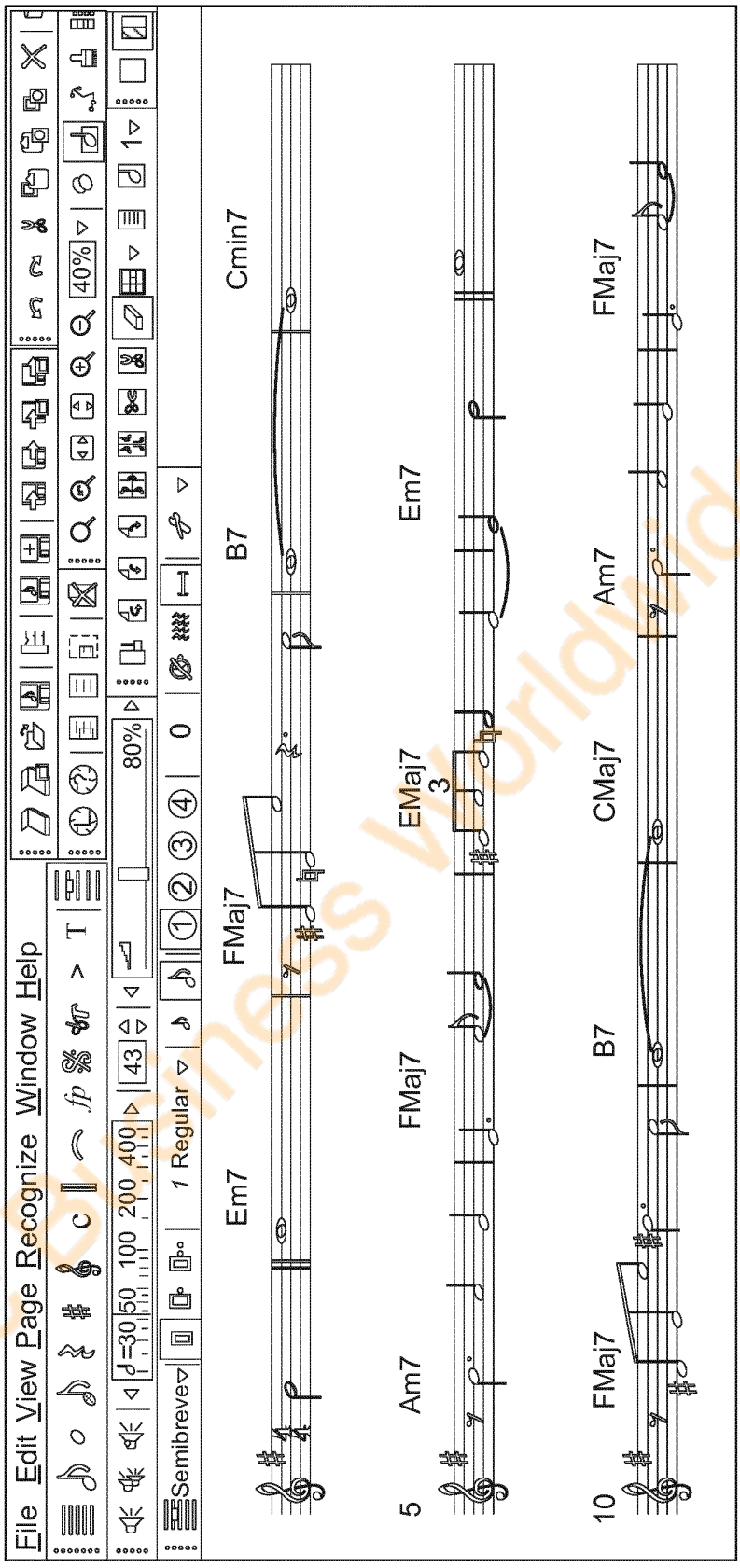


FIG. 6

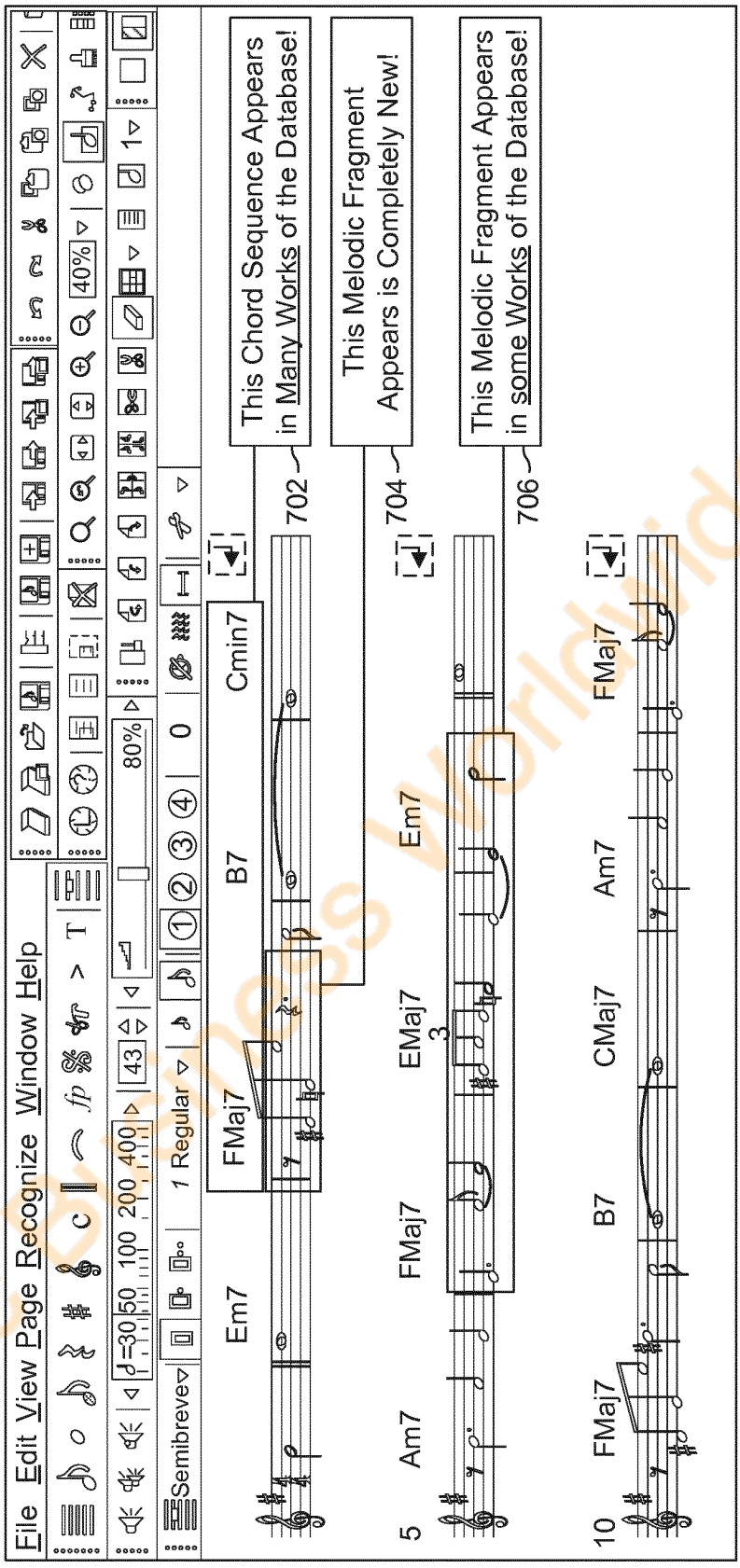


FIG. 7

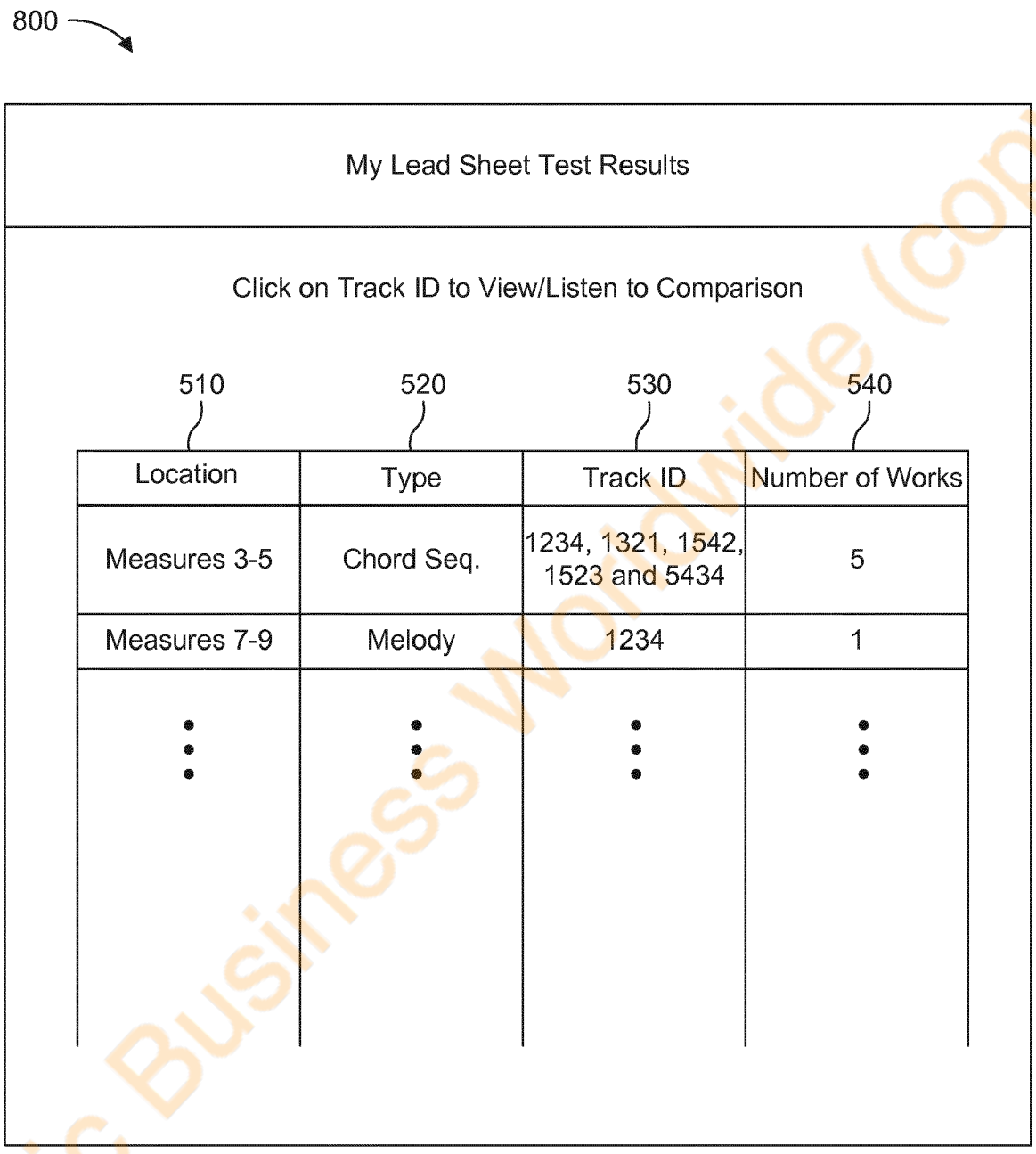


FIG. 8

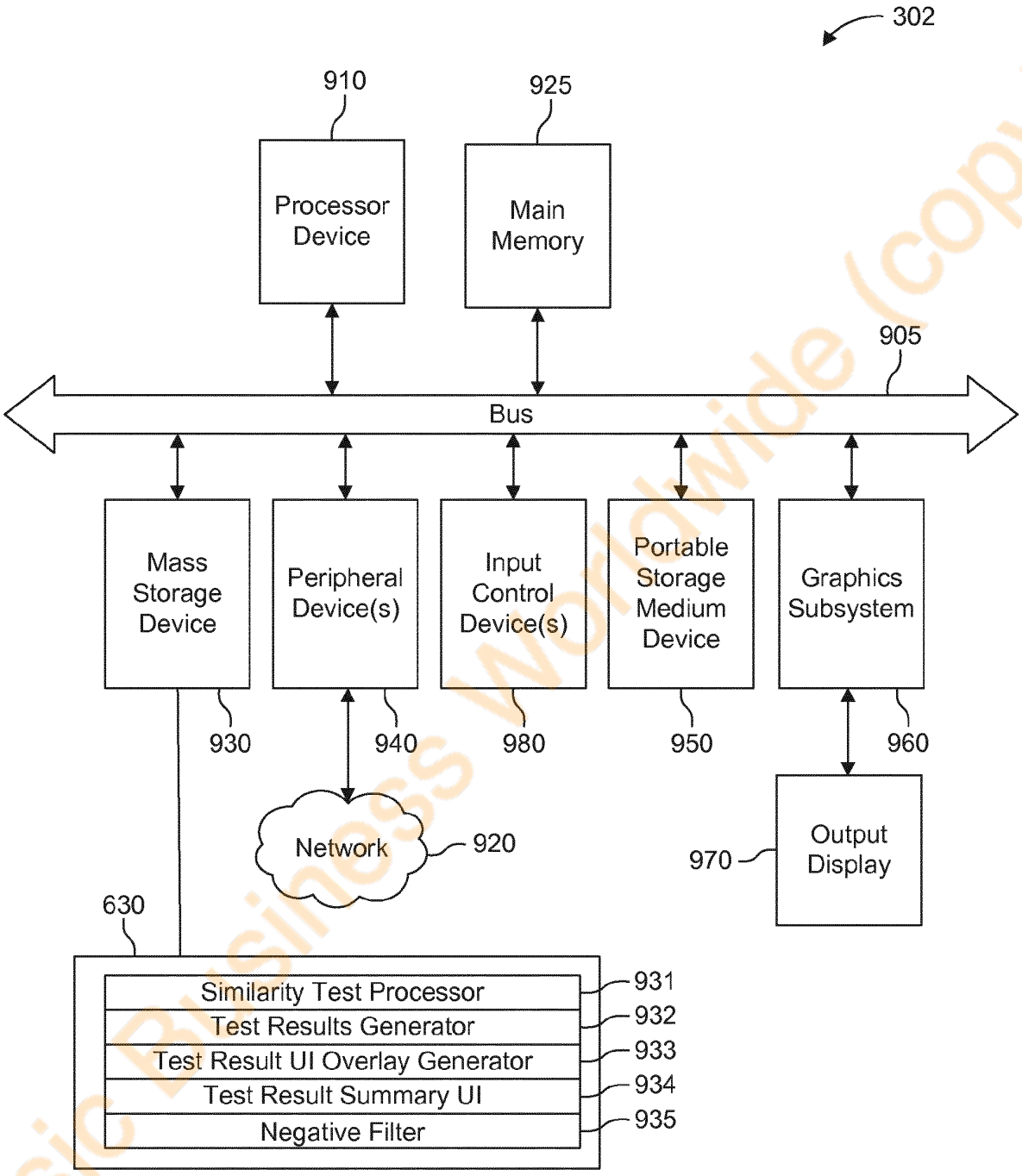


FIG. 9

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 19 17 6232

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	JP 2001 265324 A (P & P KK; YAMAKITA SUEHIRO) 28 September 2001 (2001-09-28) * abstract; figures 1-20 * * paragraph [0001] - paragraph [0007] * * paragraph [0050] - paragraph [0076] * * paragraph [0100] - paragraph [0131] *	1-15	INV. G10H1/00
X	US 2015/317965 A1 (HORVATH RALPH [US]) 5 November 2015 (2015-11-05) * abstract; figures 1-16 * * paragraph [0005] - paragraph [0007] * * paragraph [0058] - paragraph [0109] * * paragraph [0131] - paragraph [0141] *	1,8,15	
X	DE PRISCO ROBERTO ET AL: "Visualization of Music Plagiarism: Analysis and Evaluation", 2016 20TH INTERNATIONAL CONFERENCE INFORMATION VISUALISATION (IV), IEEE, 19 July 2016 (2016-07-19), pages 177-182, XP032955346, DOI: 10.1109/IV.2016.56 [retrieved on 2016-08-31] * abstract; figures 1-2; table 1 * * Sections 3-4 *	1,8,15	TECHNICAL FIELDS SEARCHED (IPC) G10H G06F G10G
X	JEONG-IL PARK ET AL: "Music Plagiarism Detection Using Melody Databases", 17 August 2005 (2005-08-17), KNOWLEDGE-BASED INTELLIGENT INFORMATION AND ENGINEERING SYSTEMS; [LECTURE NOTES IN COMPUTER SCIENCE;LECTURE NOTES IN ARTIFICIAL INTELLIGENCE;LNCS], SPRINGER-VERLAG, BERLIN/HEIDELBERG, PAGE(S) 684 - 693, XP019015766, ISBN: 978-3-540-28896-1 * abstract * * sections 2-5 *	1,9,15	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	11 November 2019	Lecointe, Michael

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 19 17 6232

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

11-11-2019

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
JP 2001265324 A	28-09-2001	JP 3454307 B2	06-10-2003
		JP 2001265324 A	28-09-2001
US 2015317965 A1	05-11-2015	US 2015317965 A1	05-11-2015
		US 2016098978 A1	07-04-2016

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **LEE, J. et al.** Music Plagiarism Detection System. *ITC-CSCC,* 2011 **[0005]**
- **CHRISTIAN DITTMAR et al.** *Audio Forensics Meets Music Information Retrieval - A Toolbox For Inspection Of Music Plagiarism,* 2012 **[0006]**