IN THE INTERNATIONAL SEARCHING AUTHORITY

International Application No.	PCT/US2017/016756
International Filing Date:	6 February 2017
Priority Date:	5 February 2016
Applicant	New Resonance, LLC
Application Title:	MAPPING CHARACTERISTICS OF MUSIC INTO A VISUAL DISPLAY
Agent Reference No.:	104058-0014/003WO1

REQUEST FOR RECTIFICATION OF OBVIOUS MISTAKES UNDER PCT RULE 91

International Searching Authority
European Patent Office
P.B. 5818 Patentlaan 2
NL-2280 HV Rijswijk

Dear Sirs,

Applicant hereby requests rectification of the below-listed obvious mistakes in the above-referenced PCT application. Accordingly, please replace pages 8, 20, 21, 26, 40, 42, 43, 44, 47, 48, 56 and 62 of the International Application, as filed, with the attached replacement pages, evidencing amendments to the specification, which replacement sheets show the addition of language, as evidenced by underlined text.

Each of the proposed amendments addresses an obvious mistake in the application as follows:

On page 8, we wish to introduce an explanation for the abbreviation "PACO", which abbreviation is otherwise not found until page 48. This item introduces no new matter but serves to clarify the text for the reader.

Pages 20, 21, 44, and 47 originally referenced "Appendix A" where "Appendix B" was intended. Applicant wishes to reference the correct appendix.

Page 26 originally referenced "the Appendix" where "Appendix B" was intended. Applicant wishes to reference the correct appendix.

Page 40 originally referenced "Appendix B" (twice) where "Appendix A" was intended in both instances. Applicant wishes to reference the correct appendix.

On page 42, Applicant wishes to insert "and an ambience score", to bring the text into conformity with reference numeral 203 in FIG. 8, to which the text refers.

On page 43, Applicant wishes to insert "and calculating a time streaming ambience score from individual ambience scores taken from successive TSX segments" to bring the text into conformity with reference numeral 206 in FIG. 8, to which the text refers.

On page 43, Applicant wishes to insert "time-streaming ambience score" to bring the text into conformity with the remaining description of FIG. 8.

Page 48 originally referenced "the Appendix" where "Appendix A" was intended. Applicant wishes to reference the correct appendix.

Page 62 originally referenced "Figure X" where "Figure 11" was intended. Applicant wishes to reference the correct figure.

On page 56, Applicant wishes to delete the reference numerals "926" and "930", which are otherwise not found in FIG. 12 and are therefore superfluous.

Please consider and integrate these amendments accordingly. This request for rectification of obvious mistakes is timely under PCT Rule 91.2, being submitted before 26 months from priority have elapsed.

Respectfully submitted,

/Richard G. A. Bone/

Date: December 28, 2017

Richard G. A. Bone

Registration No. 56,637

Correspondence Address
McDermott Will & Emery
500 North Capitol Street, N.W.
Washington, D.C. 20001-1531
USA

separately. The display is adaptive, which means that the technology can monitor and adjust the display as the music varies in complexity, or can provide a consumer with options to control adjustments on the display. Alternately, the technology provides a producer-adjustable display so that music producers and concert organizers can generate a "PACO Track" (a saved visual display from a given piece of music and PACO means "psychoacoustic color organ") that can be played and replayed.

[0025] The technology can be developed with a suite of as few standardized mappings as are effective, to enhance consumer learning and consumer ability to make use of displays mapped from audio to visual at a high level of information and detail. Yet also the technology is capable of applying a very large set of alternative mapping systems, to provide highly compelling displays tuned specially to each piece of music.

[0026] The fact that there is a structured, systematically developed set of visual cue vocabularies means that the technology is versatile and adaptable and applicable to any form of music, regardless of genre, and including sound sources such as mechanical sounds that humans would not necessarily categorize as music. The fact that the technology is equipped with applications of machine learning and Bayesian inference to improve pattern recognition, and to improve the ability of the device and the user to select the most effective mapping means that the technology can continually improve.

[0027] Other musical uses of the technology include rehearsal aids for performers, music training aids, a system for taxonomizing musical pieces, such as for search and retrieval from digital repositories, and providing a platform for psychoacoustic research.

[0028] Additional objects, advantages, aspects, and novel features of the invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon reading the following, or may be learned by practice of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] FIG. 1 is a flow chart illustrating three stages involved in the conversion of a music source file to a corresponding dynamic visual representation of the music, as further described herein.

presence or absence of columns and columns dividing sections of music, e.g., by timbre or melody-harmony-percussion lines; presence or absence of horizontal bands dividing sections of the music, again e.g., by timbre or melody-harmony-percussion lines, the width of any such columns, subcolumns, and horizontal bands; color, patterns or textures in horizontal bars at the top or bottom of the display, those colors, patterns or textures depicting characteristics of musical phrases such as chord progression, affect and tension, aligned or not in time and/or pitch with the notes having those characteristics; background color and color changes, with color and color intensity optionally varying spatially on a display, and/or with time; and blending or distinctness of two or more visual cues. For any of the listed cues involving color, that color can vary in hue, saturation, iridescence or shimmer. Any of the listed cues can include a gradient over whatever spatial extent is involved. Any of the listed cues can characterize a region of the display, including a frame around the display or around a part of the display. Any of the listed cues can be varied ordinally to indicate an ordinal variation in the audio cue being represented; that ordinal variation can vary as a monotonic, linear, ratio or logarithmic function of the ordinal variation in the represented audio cue; where appropriate that visual cue ordinal variation can be scaled to the magnitude of the effect in the represented audio cue. A single visual cue may also contain two or more component parts, such as a note icon and a visual cue modifying that note icon, e.g., an instrument symbol within, attached to, or adjacent to the note icon. ~~Appendix A~~ Appendix B presents a visual cue vocabulary for each audio cue considered here, i.e., listed in the following.

[0081] Visual cues for the following audio cues are of particular importance in the context of the present technology: a note, which can be characterized by any of the audio cues listed as follows: pitch; amplitude; time of note onset; note duration; pitch interval between at least two simultaneously played notes, thus including both interval and chord; pitch interval between a note and a previous note; extending that to an arpeggio; different singers and/or instruments as represented by timbre; the number of a particular instrument or the number of particular voices creating a given note; extending that to numbers of different instruments and/or voices creating a given note; sibilance; attack and decay of a note, strum or chord; strum; melody versus harmony versus percussion line; transitional note; overall volume of a musical piece; chord progression; affect; tension; ambience;

vibrato; tremolo; and glissando. ~~Appendix A~~ Appendix B describes each of those audio cues in further detail.

[0082] Time streaming is one aspect of the consumer's sense of perceptual conformality, insofar as in time streaming, note icons appearing at one or more points and/or lines on the display stream until they vanish at one or more points or lines on the display. That is perceptually conformal with the consumer's audio experience, where a note appears at a particular point in time, the current time, then persists in his/her memory over some period of time moving on through that period of time, then vanishes from his immediate memory.

[0083] Visual cues may also include text corresponding to the lyrics of a song, either determined by the system or, more typically, provided as part of the music source file. In some instances lyrics can be ascertained from vocal music using commercially available speech recognition software.

[0084] The amplitude and timbre of each note are central to music appreciation, and so should be depicted as directly, clearly and completely as possible.

[0085] There are note modifiers that are quite important to music appreciation that should be depicted, while they are not depicted in typical MIDI formats. Those modifiers include attack, sibilance, multiple instruments playing the same note, and tremolo.

[0086] It must be emphasized that the foregoing audio cues and visual cues are for purposes of illustration, and those described herein are not intended to be an exhaustive list.

[0087] The number of cues selected for mapping into the visual display is preferably optimized to provide the best user experience. This is essentially a matter of visual bandwidth management, where the term "bandwidth" is used herein as referring to information displayed per second in the particular format chosen. Visual bandwidth management is important herein insofar as there are limits to human visual perception, and the present method and system should operate within that perceptual bandwidth.

Alternatively, or in addition, the notes within the aforementioned musical lines may be highlighted or otherwise identified on the display.

[0099] In some embodiments, perceptual conformality involves one-to-one correspondence between a group of two or more selected audio cues that are perceptually associated with each other and a group of the selected two or more corresponding visual cues. That is, in this embodiment, perceptually conformal one-to-one correspondence requires that a group of two or more perceptually associated audio cues be translated into a corresponding group of two or more perceptually associated visual cues. It will be appreciated, in this embodiment, that two audio cues that are not perceptually associated with each other are represented by two spatially separate visual cues.

[0100] A representative mapping of selected psychoacoustic cues to corresponding visual cues using a perceptually conformal mapping system is as follows. Each psychoacoustic cue is associated with an individual note or a group of notes at any one point in time or as a sequence. The existence of an individual note can be represented in the selected visual display by a square, rectangle, circle, diamond, triangle, "roundtangle" (a rectangle with rounded corners), mouth, guitar or other instrument, or other visual cue as described elsewhere herein. The shapes can optionally be borderless. That representation will be referred to herein as a note icon. There is perceptually conformal one-to-one correspondence between the auditory perception of each note and the visual perception of each corresponding visual cue. The pitches of the notes processed by the system into visual cues are not limited to the discrete notes within a standard piano keyboard, such as the 88- note or 97-note versions, but can include pitches of notes in between those discrete notes. This is particularly useful, for instance, in representing glissando, vibrato, portamento, and the like. Accordingly, described elsewhere herein – in ~~the Appendix~~ Appendix B– are audio cues with corresponding representative visual cues and a brief indication of how perceptual conformality is achieved.

Three Stages of a Method of Mapping Musical Characteristics into a Visual Display

[0101] The method of visualizing music is a three-stage process, schematically illustrated in FIG. 1. In a first stage, a music source file 100 is translated into a time stream of music data files. In the second stage, the stream of music data files is converted into a time

probability of each LIV given the GOF scores, and updates the thresholds used to convert the Bayesian Inference results into decisions. That experience is accumulated from full-note-duration analysis in three forms: (1) during a musical piece, including LIVs that stop and then start again later in the piece; (2) from past plays in the consumer's play set, for enhanced identification in later plays; and (3) new notes, for possible use in later plays in the consumer's play set.

[0152] Those improvements in performance fall into three categories: 1) More refined patterns for a more refined set of LIVs to be detected and identified, based on logging the observed patterns, *i.e.*, if different patterns are logged for female soprano voices, those can be identified as different LIVs and labeled accordingly, perhaps to be matched to named performers through external-source downloads. The same process applies to, *e.g.*, more effectively and rapidly distinguishing viola from violin; 2) That same more refined pattern recognition process, but in particular applied to learning a LIV early in a piece then identifying that LIV when and if it reappears in that piece; and learning a LIV from a user's set of played music, then identifying it more quickly when and if it reappears in other played music; 3) More rapid identification of LIVs, based on the inference sequences that eventually result in a LIV identification.

[0153] ~~Appendix B~~ Appendix A presents a very general set of alternative mappings from audio cues to visual cues. That set of mappings is general enough to provide a visual cue vocabulary that can effectively support the broad range of implementations of the device described in ~~Appendix B~~ Appendix A and FIG. 11. The dimensions of that broad range of implementations can be summarized in 6 aspects:

- Aspects 1 (Source Complexity) and 2 (Genre) describe the music to be mapped.
- Aspects 3 (Implementation Mode and so Signal Processing Power Called For) and 4 (Display) describe the technical aspects of the implementation.
- Aspects 5 (User Experience and Needs) and 6 (User Preference) describe the user aspects of the implementation.

Stage 2, and in addition it may be provided as a separate output, 124. That separate output may be provided to a consumer to be used in connection with a user-selected, separately acquired Stage 2 device. Formats of output 124 can take other forms as well, e.g., as a digital music file (such as a MIDI file) or a musical score.

Stage 2

[0157] The second stage of the method is illustrated in FIG. 8. The purpose of the second stage is to take the output of Stage 1, i.e., a time stream of PAL tables 123, and convert it to a time stream of psychoacoustic attribute files, PAFs, 214. The format of a representative PAF is presented in FIG. 9.

[0158] In overview, Stage 2 analyzes the time stream of PALs to extract all the remaining cues to be used by the system, all cues other than pitch, amplitude and LIV. The input time stream of PALs can originate within the device (as 123), or from a different device separately acquired by the listener (as 125) such as a separately acquired MIDI file, or a Stage 1 output from a different system. Stage 2 is comprised of five levels, as follows:

Stage 2, level 1

[0159] The first level of Stage 2 calculates across-all-note, within-TSX, metrics, of which there can be the following four, among others: 201: summing the amplitudes of all notes in a particular TSX to give a total TSX amplitude or volume; 202: calculating one or more chordal structures from frequency ratios; 203: assigning an individual affect score (i.e., an affect score for one TSX) based on factors such as pitch, tempo, key (*i.e.*, major or minor) and instrumentation, and an ambience score; and 204: assigning an individual tension score (i.e., a tension score for one TSX) based on several factors, including such as chord inversions and chord progressions, intervals, relationships between melody and harmony lines, relationships between multiple melody lines, relationships between current notes and the tonic, and volume. Each of the foregoing metrics is calculated for a single time segment TSX (*cf.* FIG. 8). The methods herein are not limited to those four metrics.

[0160] Processing in level 1 of stage 2 thus provides summed amplitudes, calculated chord structures, affect scores, ambience scores, and tension scores for each TSX analyzed. Other calculations in Stage 2, *i.e.*, the calculations for levels 2 through 4, involve

analysis of not only the current TSX but also a plurality of preceding TSXs. The number of preceding TSXs analyzed depends on the particular metric provided, as will be further described.

Stage 2, level 2

[0161] The second level of stage 2 calculates across-many-TSX metrics of the musical piece, with four individual calculations performed using information obtained in level 1 of stage 2 for the current and preceding TSX segments, as follows. 205: Calculating a chord progression metric by using pattern recognition of the chord structure metric across successive TSX segments; 206: Calculating a time-streaming affect score from individual affect scores taken from successive TSX segments and calculating a time-streaming ambience score from individual ambience scores taken from successive TSX segments; 207: Deducing a tonic using an algorithm that reviews multiple TSX segments; and 208: Assigning a time-streaming tension score by combining individual tension scores obtained in 204 with the tonic identified in 207.

[0162] The metrics provided by the foregoing calculations, chord progression, time-streaming affect score, tonic, time-streaming ambience score and time-stream tension score, are calculated based on a sequence of preceding TSX segments through the current TSX, as noted above. The number of preceding TSX segments analyzed can vary with the metric calculated, such that n1 represents the number of TSX segments required to calculate chord progression, n2 the number required to calculate affect score, n3 the number required to deduce a tonic, and n4 the number required to assign a tension score. Each individual n value may be different for different musical pieces and/or different types of musical pieces. For instance, n4, the number of TSX segments required to assign tension score, will be much greater for a complex orchestral piece but much smaller for a short piano piece that is simple in structure. In fact, n4 can be adaptive to the musical piece as it progresses, as it is analyzed over time.

Stage 2, Level 3:

[0163] In the third level of stage 2, three note-oriented metrics are calculated over many TSX segments: attack 209, strum 210, and assignment to a melody or harmony line 211, if appropriate. In Level 3, each metric pertains to a single note. The attack of a note is identified by pattern recognition of its amplitude onset; the pattern of note onset includes speed of onset. Notes are identified as contained within a strum by pattern recognition of a

rapid note series. For assignment of a note to a melody or harmony line, if appropriate, pattern recognition is based on a series of notes all having the same LIV, for example, a sequence of notes played by a violin, sung by a female voice, and the like. A note is typically assigned to a melody line if it is contained within a sequence of same LIV notes where that sequence fits into a typical melody pattern that can be inferred from relative pitch, relative amplitude, and, for a mix of voice and instruments, voice. The same reasoning is true for identifying harmony.

[0164] As in level 2 of stage 2, the number of TSX segments analyzed may be different for each metric, such that n5 represents the number of TSX segments required to calculate the attack pattern of a note, n6 is the number required to determine the presence of strum, and n7 the number required to assign a note to a melody or harmony line. It will be appreciated that n7 will typically be much higher than n5 and n6 since the melody and sometimes harmony may only become apparent over one to several seconds. As in level 2 of stage 2, n7 can be adapted to the musical piece as it progresses, as it is analyzed over time.

Stage 2, Level 4

[0165] As with level 3, the metrics determined in level 4 of stage 2 are note-oriented, 212. In level 4, a nine-element vector is created that characterizes each note with the following information: (1) the status of the note in each TSX, *i.e.*, as beginning, continuing, or ending; (2) the pitch of the note; (3) the amplitude of the note; (4) the assigned LIV from the PAL data set; (5) N-instrument; (6) sibilance; (7) attack; (8) strum; and (9) melody/harmony/neither (characterization of a note as within a melody, within harmony, or neither). It is to be understood that the foregoing 9 elements are not the only ones that can be used to create a vector to characterize a note; other elements can be used in addition to, or in place of, those 9. Additionally, a satisfactory vector can be created with smaller numbers of elements, such as 6, 7, or 8.

[0166] There are other attributes that correspond to notes extending throughout a sequence of TSX segments that will be visually apparent solely from amplitude and frequency mapping of TSX data. These are tremolo, vibrato, and glissando, though as noted in ~~Appendix A~~ Appendix B, those audio cues can also be enhanced by special visual cues.

[0174] The operations of Stages 1, 2, and 3 are performed by a device having any one of a variety of configurations. For instance, the device can be a digital signal processing circuit inside in a consumer entertainment device or packaged in a separate housing. It can also be implemented in music processing devices for music producers and concert producers.

[0175] FIG. 10 schematically illustrates a representative display showing possible visual cues accompanying a segment of music; this example of a display shows how the visual cues described previously can be displayed. The circled numbers in FIG. 10 correspond to the cue numbering elsewhere herein (see ~~Appendix A~~ Appendix B, for example).

[0176] The system can store information associated with a piece of music it processes, such that the music is stored along with the set of audio cues identified. Over time, the stored information can grow, e.g. at a market-wide scale, and ultimately be used as reference library that the system can query to find a particular piece of music or type of music. For instance, a consumer may wish to find a piece of music in a particular key with a particular affect played by a particular instrument, and can query the stored information in order to identify such a piece of music.

Applications

[0177] The method has several applications that do not depend on the real-time performance of a full implementation of all of the operations described herein. The system can be modified in one or more ways to reduce its overall computational burden, so that it may be made available to a variety of end users with different needs and/or expectations. Such modifications include, without limitation: capability of operating in real time, *i.e.*, capability of processing as music is presented; operation at different levels of time resolution; sophistication of mapping; level of detail in voice/singer identification (*e.g.*, female, generically, versus specific individual such as Taylor Swift or Marilyn Horne); level of detail in instrument identification (e.g., string instrument versus viola); sophistication of melody-harmony recognition; and sophistication of options offered to the user on a control device.

[0178] The various types of user can be placed in three categories: individual consumers; concerts; and commercial music producers. These applications are also

discussed in ~~the Appendix~~ Appendix A, where they are discussed from the perspective of their implications for called-for signal processing power. In the following, applications are discussed from the perspective of the implications of those markets for the intrinsically robust value of the device.

[0179] When targeting individual consumers, the device can be implemented at any of several price points. If it proves to be too expensive for some consumers to include real-time performance at a universally attractive price, then the system can be implemented in higher-cost versions for real-time performance, but also in lower-cost versions that provide simplified performance in real time, and/or in a two-pass mode, where the system can accept a music file and analyze it over an extended period of time, then store its analyzed file for playback synchronized with the music at any later time chosen by the consumer. The two-pass mode offers the opportunity for enhanced performance by allowing the system to preview the entire piece, and make adjustments regarding amplitude range, pitch range, LIV identification, melody-harmony divisions, chord progressions, affects and tensions, where those adjustments can only be made less effectively in a real-time mode.

[0180] Concert performances can preferentially employ a high performance version of the method and system so as to generate high performance in real time. In addition, there are several aspects of concerts that make the music cue extraction tasks much easier: LIV identification can be fully accomplished simply by separate microphone connections, including the specification of N voices or musical instruments versus single ones; melody-harmony divisions can be specified by a combination of microphone connections and real-time manual adjustments; chord progressions, affects and tensions can be specified by algorithms but also supplemented by real-time manual adjustments; and amplitude ranges and pitch ranges can be set in rehearsal. In an alternative embodiment, different instruments playing together may each be operatively connected to their own system, each including a separate display, such that each instrument's music is visualized simultaneously on different displays. In addition, concert producers can broadcast a PACO Track signal to the audience members' personal mobile devices (with PACO referring to psychoacoustic color organ). That PACO Track can either allow the audience member to select among

overcrowding the display with too many visual cues is higher. A lower resolution display, similarly, will typically call for fewer visual cues.

[0205] The computing devices can have suitably configured processors, including, without limitation, graphics processors, vector processors, and math coprocessors, for running software that carries out the methods herein. In addition, certain computing functions are typically distributed across more than one computer so that, for example, one computer accepts input and instructions, and a second or additional computers receive the instructions via a network connection and carry out the processing at a remote location, and optionally communicate results or output back to the first computer.

[0206] Control of the computing apparatuses can be via a user interface, which may comprise a display 924, mouse ~~926~~, keyboard ~~930~~, and/or other items not shown in FIG. 12, such as a track-pad, track-ball, touch-screen, stylus, speech-recognition, gesture-recognition technology, or other input such as based on a user's eye-movement, or any subcombination or combination of inputs thereof. Additionally, implementations are configured that permit a user to access computer 900 remotely, over a network connection, and to view the visual depiction of music via an interface having attributes comparable to display 924. The interface there for may comprise a microphone input for accepting musical sounds for processing. Music, in the form of a data file, may also be introduced into computer 900 via network interface 936 as well as via a plug-in memory stick or other media.

[0207] In one embodiment, the computing apparatus can be configured to restrict user access, such as by scanning a QR-code, or requiring gesture recognition, biometric data input, or password input before the visual display is started.

[0208] The manner of operation of the technology, when reduced to an embodiment as one or more software modules, functions, or subroutines, can be in a batch-mode – as on a stored database of audio data, processed in batches, or by interaction with a user who inputs specific instructions for a single piece of music.

[0209] The results of converting audio data to visual form, as created by the technology herein, can be displayed in tangible form, such as on one or more computer displays, such

set to a user's experience level, he may want to choose more or less complex displays and so mappings, from the maximally complex to much less complex, more "abstract" displays and so mappings. The user may want to increase his skill level, and so temporarily set the device to above his experience level. Settings could include for example "Abstract," "Party," and "Maximal."

[0231] Reviewing the Six Aspects: Aspects 1 through 4 can be set to their values by the device system, while Aspect 5 can be set to its value by a combination of user-specific system experience and user input, while Aspect 6 can be set to its level by the user.

Implications of Those Six Aspects

[0232] The ability of the device to select from a range of mappings is critical to its functioning in the range of implementations defined by the six aspects. The relationships among the six aspects of implementation and the mapping employed by the device are presented in FIG. 11.

APPENDIX B

The Mapping System

Requirements for the Mapping System

[0233] The Mapping System presented here was developed to meet two requirements: Requirement 1: must provide the opportunity to make the most effective use of the visual perception space. Requirement 2: must provide the opportunity to flexibly and effectively accommodate all plausible sets of values of the six aspects of Appendix A and ~~Figure X~~ Figure 11.

[0234] To that end, requirement 1 can be subdivided into two parts. In Requirement 1a, all visual cues are selected from a very broad vocabulary of cues, described in terms of six categories of cue descriptors, listed herein. This is another aspect of the bandwidth concept described elsewhere herein. Again, that bandwidth is not in terms of physical bits per second – it is in terms of what the user can perceive and comprehend in the display. So the vocabulary of visual cues used here is designed to exploit patterns of human visual perception, in particular: spatial position and extent, at the scale of lines, icons (with or without borders), bands or regions of the display including the borders, sizes and shapes of those icons, bands and regions, and their color (hue, saturation, shimmer and iridescence)