



(11) **EP 3 579 223 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**11.12.2019 Bulletin 2019/50**

(51) Int Cl.:  
**G10G 1/00 (2006.01)**      **G10H 1/00 (2006.01)**  
**G10H 1/40 (2006.01)**      **G09B 15/02 (2006.01)**

(21) Application number: **18382392.1**

(22) Date of filing: **04.06.2018**

(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**

(72) Inventors:  
• **ESTENOZ ABENDAÑO, Ainhoa**  
**31195 Larrageta - NAVARRA (ES)**  
• **ZIVANOVIC JEREMIC, Miroslav**  
**31195 Larrageta - NAVARRA (ES)**

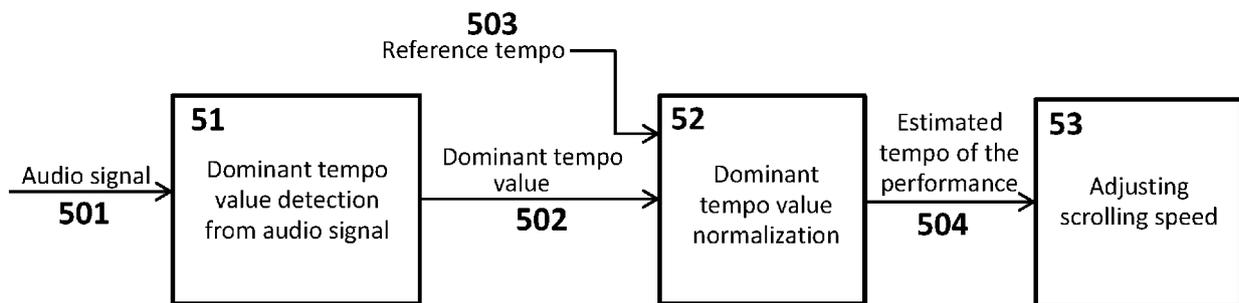
(74) Representative: **Balder IP Law, S.L.**  
**Paseo de la Castellana 93**  
**5ª planta**  
**28046 Madrid (ES)**

(71) Applicant: **NewMusicNow, S.L.**  
**31195 Larrageta, Navarra (ES)**

(54) **METHOD, DEVICE AND COMPUTER PROGRAM PRODUCT FOR SCROLLING A MUSICAL SCORE**

(57) A method for displaying a musical score on a screen of an electronic device, comprising: loading a file having a digital score in a piece of memory of said electronic device; scrolling said digital score on the screen of the electronic device; capturing an audio signal (501) corresponding to said musical score being played by a musician; repeatedly selecting frames (frame\_i) of the captured audio signal (501) and, for each selected frame (frame\_i): obtaining a dominant tempo value (502) at which the music contained in said frame (frame\_i) is

played; from the dominant tempo value (502) obtained from said frame (frame\_i) and from a reference tempo (503) comprising a reference figure and a reference tempo value, estimating (52) the tempo at which the musician is playing, said estimated tempo (504) comprising said reference figure and a normalized tempo value with respect to the dominant tempo value (502); adjusting the scrolling speed of the digital score according to the estimated tempo (504).



**FIG. 5**

**EP 3 579 223 A1**

**Description****TECHNICAL FIELD**

5 [0001] The present invention relates to the field of digital signal treatment and, in particular, to the treatment of digital signals corresponding to a musical performance and to methods and systems for recognizing music as it is performed. The invention also relates to methods and systems for displaying and scrolling musical scores on a display screen.

**STATE OF THE ART**

10 [0002] It is well-known that musicians sometimes use virtual scores or electronic scores, rather than conventional paper ones. Electronic scores have meant, among other advantages, appreciable savings in terms of paper and, as a consequence, space.

15 [0003] One of the main challenges faced by electronic scores has been how to scroll or advance virtual sheet music for a user playing an instrument. Current solutions for displaying electronic scores are based on a page-by-page system. This means that electronic scores are stored page by page on a storage medium. When a page has been shown for a certain time period, that page disappears from the screen and the following page is shown. At the end of the time period, the portion -that is to say, page or slide- of musical notes that is shown on the screen is automatically replaced with a subsequent portion -such as page or slide- of musical notes, which again stays on the screen for a certain time period.

20 For example, United States patent US7098392B2 discloses a method for providing for video display of music responsive to the music data stored in a music database. In this method, first, a page of music image data from a music database is defined; next, ordered logical sections within that page are defined; then, the mapping is stored in a memory for selective retrieval; finally, the video display of the music responsive to the mapping and the storing is provided.

25 [0004] European patent EP2919228B1 discloses a method for scrolling a musical score on a screen of a device, in which musical signs are scrolled on the screen by continuously showing on the screen additional signs of music while the already scrolled ones disappear from the screen. The scrolling speed is adjusted according to the music content being displayed on the screen.

30 [0005] The problem of how quickly or slowly to scroll or advance virtual scores for a user playing an instrument, has already been addressed. In fact, different attempts have been made in the last years for analyzing in real-time audio signals resulting from a performance of a piece of music and tracking the corresponding location in the musical score of the piece.

35 [0006] United States patent US8530735B2 describes a method for displaying music on a display screen, in which a tempo of the user's performance is supposed to be detected, from which the time period required by the player to complete the performance of a displayed portion of musical notes is calculated. At the end of the calculated time period, the portion of musical notes displayed on the screen is automatically replaced with a subsequent portion of musical notes. In this respect, M.F. McKinney et al. in Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms, Journal of New Music Research, 2007, Vol. 36, No 1, pp. 1-16, provide an extended analysis of eight different algorithms for musical tempo extraction and beat tracking. While obtaining the tempo of a musical record has been successfully achieved for recorded musical pieces, extrapolating current methods thereof to real-time live performances has been

40 proved to be unsuccessful due to noise and other disturbances.

[0007] Another example of these attempts is disclosed in US6156964, which refers to a method of displaying a musical score in which a portion of the music score data corresponding to the playing position of the musician is displayed on a display device. The playing position of the musician in order to display the appropriate portion of the score on the screen is determined by comparing tone frequency data of the music score with tone frequency data of the music being played.

45 Another attempt for performing real-time music note recognition is disclosed in US2005/0015258A1, in which a played note is identified and compared with a reference note by identifying the starting and ending edges in the time domain of each note. Other proposals for display scrolling based on determining the notes associated with an input signal captured by a microphone are disclosed in US9280960B1, US9747876B1 and US2001/0023635A1. A disadvantage of these approaches is that many unexpected aspects may have an influence on the tone frequency of a same type of instrument or even on the same instrument, such as room temperature or tuning. Besides, unequivocally identifying a note is not feasible in real time, because notes are composed of harmonics and therefore only in ideal circumstances (in the absence of noise and wherein instruments generate very clear signals) could notes be clearly recognized. This is illustrated in Figure 1, which shows on the left respective typical signals captured when a guitar (up) and a clarinet (down) are individually playing. As can be observed, while guitar notes can be identified, clarinet notes cannot. In general,

50 depending on the instrument, notes are generated in a different way: hard onsets and soft onsets. Hard onsets are typically produced, for example, by stringed instruments, such as guitar. Soft onsets are typically produced, for example, by wind instruments, such as clarinet, in which a change in note is not clearly defined. In Figure 1, on the right, a graphic showing a signal representing the combined captured music played simultaneously by a guitar and clarinet is depicted.

55

It is impossible to identify which notes are being played. Last but not least, in US2005/0015258A1 a training database is required, in such a way that, for a given instrument, a calibration procedure needs to be performed, that identifies the key features of each note in a range of notes and stores them in a pattern database.

5 [0008] A different attempt is disclosed in US8660678B1, which refers to a method for following a score based on Markov chains. Instead of focussing on analysing the detected audio signal, probabilistic technics are used, both for reducing the processing workload and for trying to avoid the problems derived from audio signals analysis. The most likely current location in the score and the most likely current tempo are estimated. However, again, this method cannot unequivocally identify the music played by any instrument and requires training beforehand a software application in order to generate Markov models.

10 [0009] In sum, well-known efforts for music scrolling based on real-time estimation of the music being played by a musician have pointed out to be unfeasible for the reasons enumerated above.

## DESCRIPTION OF THE INVENTION

15 [0010] The present disclosure provides a method for scrolling a digital musical score on a screen of an electronic device based on real-time music recognition which overcomes the mentioned disadvantages. The scrolling speed is adjusted in real time according to the real time tempo at which music is being played by a musician.

20 [0011] The method described herein is mainly designed to run on an electronic device, such as a personal digital assistant (PDA), a portable reader device, a tablet, a cell phone containing a display or any device comprising a memory, a processor and a screen or display. An audio sensing means, such as a sound sensing means or a vibration sensing means, is also required in order to capture the sound produced by the one or more musicians. The sound may be produced by traditional instrument(s) or digital one(s), such as a MIDI board. The sensing means may be embedded in the electronic device or may be a separate device connected to the electronic device by means of a wired or wireless connection. Non-limiting examples of audio sensing means are a microphone or any other sound or vibration capturing means, such as piezo-electric capturing means. In some embodiments of this disclosure, the term "audio signal" refers to the signal as captured by the audio sensing means. The capture takes place in real time, that is to say, as music is being played. The captured signal is typically an analog signal. While it is captured, it may be converted into a digital signal. In some embodiments of this disclosure, the term "audio signal" may refer to the already digitized analog signal by means of an A/D converter (analog-to-digital converter), preferably embedded in the electronic device.

25 [0012] The audio sensing means may be comprised in the electronic device or may be independent therefrom, in which case the audio sensing means is connected to the electronic device. As will be apparent in the light of this disclosure, the execution of the current method does not require high computational workload, as a consequence of which the method is especially indicated for being executed on low and mid-range electronic devices, such as a personal digital assistant (PDA), a portable reader device, a tablet or a cell phone. The method is preferably implemented as a software application (APP). The method may be designed to run simultaneously in a plurality of such devices, for example when an orchestra or any other group of musicians is playing together. The method is implemented as computer program instructions/code which runs on one or more of the previously mentioned devices. It also requires storage means for storing the music scores in the form of digital files. This storage can be local or distributed, such as in the cloud. Optionally, additional hardware can be used, such as pedals for hands-free operation.

30 [0013] Musical metric figures or simply musical figures are individual signs, including signs representing sounds (these signs are called "notes") and signs representing musical silence (these signs are called "rests"). Each sign (notes and rests) represents within a measure a certain time period (a period of sound or a period of silence, respectively). There is a relationship between the duration in time of different musical figures (notes and rests). For example, 1 Whole note (or Semibreve) = 2 Half note (or Minim) = 4 Quarter note (or Crotchet), etc. In other words, each musical sign comprises double information: sound and time duration. So, each conventional note can be divided into a certain amount of "reference musical figures". A "reference musical figure" (also referred to as "reference figure" from now on) can be any of the former notes (whole note, half note, etc.) which is taken as a reference along a score or a portion of a score. For example, if the quarter note is taken as "reference figure", then, the whole note is made of four reference figures. An empty measure sign also has time duration of a certain number of beats, indicated in the score or by the conductor or musician. In addition, musical notes and rests can be dotted in order to lengthen their duration.

35 [0014] Tempo, which is normally expressed as beats per minute (BPM or bpm), controls the rate at which the musical signs in a line -or in general, in a score- of music are played. In a digital score, tempo is defined or expressed as a "reference figure" and a "value" (in particular, in musical language, as "reference figure = value"), wherein the "value", also referred to as tempo value in this text, represents how many times that "reference figure" -or any equivalent ones- must be played in one minute. For example, if a tempo is defined as "quarter note = 50" BPM, it means that in one minute 50 quarter notes -or any equivalent figures- must be played. Although an approximated tempo is often suggested, for example using an Italian word, such as "*Andante*", in the musical score - and can vary along the score-, in many occasions it is imposed by the conductor -in the case of a group of musicians playing together- or by the player, who does not

necessarily have to follow the tempo originally suggested by the composer, or it is obtained in a different way.

**[0015]** The musical score is in a digital format for representing, understanding and/or providing musical notation, that is to say, in a format which enables to unequivocally obtain all the symbols comprised in a score. In other words, the format must be a musical notation format, such as MusicXML format or Standard MIDI File (SMF) format or MXL format, which are well-known formats for representing musical notation, unlike other digital formats, such as PDF, TIFF, JPG, BMP, EPS, PostScript or others. For example, the MusicXML format is a fully and openly documented XML-based proprietary format for representing musical notation. The MusicXML standard contains information such as title, author, number of measures, number of systems, instrument number and name, position and duration of notes, and, generally, the same information as provided by a paper score. MIDI (*Musical Instrument Digital Interface*) is a technical standard that describes a protocol, digital interface and connectors and allows a wide variety of electronic musical instruments, computers and other related devices to connect and communicate with one another. MIDI carries event messages that specify notation, pitch and velocity, control signals for parameters such as volume, vibrato, audio panning, cues, and clock signals that set and synchronize tempo between multiple devices. The Standard MIDI File (SMF) is a file format that provides a standardized way for sequences to be saved, transported, and opened in other systems. Once a score in a musical notation format, which enables to unequivocally obtain all the symbols comprised in the score, such as MusicXML, Standard MIDI File (SMF) format or MXL format, has been open at a local device (either stored locally or in the internet, for example with restricted access) the contents of the score can be drawn on the screen of the device.

**[0016]** The contents of the digital score may be adapted to the screen of the device. From now on, the term "file" refers to a file in a musical notation format comprising a musical score. The file is preferably loaded in the device and stored locally in a buffer within the memory of the device. From now on, the term "digital score" refers to a musical score in a musical notation format.

**[0017]** According to a first aspect of the present invention, there is provided a method for displaying a musical score on a screen of a device, comprising: loading a file having a digital score in a piece or part of memory of the device; scrolling the digital score on the screen of the device; capturing an audio signal corresponding to the musical score being played by a musician; repeatedly selecting frames of the captured audio signal and, for each selected frame: obtaining a dominant tempo value at which the music contained in said frame is played; from the dominant tempo value obtained from said frame and from a reference tempo comprising a reference figure and a reference tempo value, estimating the tempo at which the musician is playing, said estimated tempo comprising said reference figure and a normalized tempo value with respect to the dominant tempo value; adjusting the scrolling speed of the digital score according to the estimated tempo.

**[0018]** Thus, the scrolling speed is adjusted in real time, according to the current tempo of the user who is actually playing. Those skilled in the art will recognize that in the context of the present invention, there will always be a delay with respect to the performer (musician) in the determination of the estimated tempo. In other words, "real-time" should be understood as guaranteeing a response within certain time constraints. In this context, "real time" is understood a time comprised within a time range varying between a lower value  $V_{\min}$  and an upper value  $V_{\max}$ . The upper value  $V_{\max}$  may be in the order of seconds, such as equal to or lower than, for example, 10 seconds, or equal to or lower than 5 seconds, or equal to or lower than 2 seconds, or equal to or lower than 1 seconds. According to current technology, the lower value  $V_{\min}$  may be, in a non-limiting way, equal to or higher than  $1 \mu\text{s}$  (microsecond,  $10^{-6}$  seconds), such as equal to or higher than 0.1 ms (milliseconds,  $10^{-3}$  seconds), or equal to or higher than 1 ms, or equal to or higher than 50 ms, or equal to or higher than 100 ms.

**[0019]** In embodiments of this disclosure, the musical score may be scrolled continuously - also referred to as dynamically, that is to say, as a continuous string of musical signs, drawn on the screen vertically, horizontally or obliquely, without interruption. In embodiments of this disclosure, the musical score may be scrolled continuously or dynamically, vertically, horizontally or obliquely, with interruptions, such as soft interruptions, when required. The scrolling speed is adjusted -increased or decreased-taking into account the music being played by the musician. In some embodiments of the invention, while the score is continuously scrolled, depending on the score and/or on the musician's performance, a portion of the digital score may be stopped or interrupted for a certain time on the screen and started to be continuously scrolled again afterwards. This may happen, for example, when a portion of score has many rests. The musician is thus enabled to read music linearly, as music really is, while adapting the scrolling to the music being played. for a certain time on the screen and started to be continuously scrolled again afterwards. scrolled again afterwards. This may happen, for example, when a portion of score has many rests. The musician is thus enabled to read music linearly, as music really is, while adapting the scrolling to the music being played.

**[0020]** In some embodiments of the invention, the dominant tempo value at which the music contained in said frame is played, is obtained as follows: detecting an onset function of said frame; finding a dominant tempo value in said onset function.

**[0021]** In some embodiments of the invention, the onset function is detected as follows: obtaining a spectrogram from the captured frame; obtaining the onset function from the spectrogram.

**[0022]** In some embodiments of the invention, the dominant tempo value is obtained by applying an autocorrelation

function to the onset function.

[0023] In some embodiments of the invention, for a frame, the estimated tempo obtained from the dominant tempo value and from a reference tempo is obtained as follows: applying a set of scaling values to the obtained dominant tempo value, thus obtaining a set of scaled dominant tempo values, calculating the absolute difference between each value of the set of scaled dominant tempo values and the reference tempo value of the reference tempo, selecting the scaling value of the set of scaling values corresponding to the lowest absolute difference, multiplying the dominant tempo value by the selected scaling value, thus obtaining a normalized tempo value of the estimated tempo, the estimated tempo comprising said normalized tempo value and the reference figure of the reference tempo.

[0024] In some embodiments of the invention, the method further comprises: after calculating the absolute difference, selecting the two lowest values  $V_1$ ,  $V_2$ , wherein  $V_1$  is the lowest value of the two values and  $V_2$  is the highest value of the two values; if  $V_1/V_2 \leq R$ , wherein  $R$  is a ratio established for limiting the deviation of the current estimated tempo, the estimation is considered correct, if  $V_1/V_2 > R$ , then the estimation is considered incorrect.

[0025] In some embodiments of the invention, if the estimation for the current frame is considered incorrect, the estimated tempo for that frame is obtained from a number of previously obtained estimated tempos for corresponding previous frames.

[0026] In some embodiments of the invention, the reference tempo used for obtaining the estimated tempo is obtained as follows: The reference tempo is manually fixed by the musician, or the reference tempo is obtained from the digital score, or the reference tempo is provided by the algorithm by default, or the reference tempo is extracted from a data base.

[0027] In some embodiments of the invention, the reference tempo is obtained from the digital score as follows: as a metronome mark, as a value associated to a word included in the score, as a time signature, as indication of metric changes along the score, or as a combination of the former.

[0028] In some embodiments of the invention, the scrolling speed of said musical score on the screen is readjusted every time a new frame of the audio signal is captured.

[0029] In some embodiments of the invention, the method further comprises, prior to continuously capturing frames of said audio signal, verifying that the first played notes correspond to a starting point identified in the digital score.

[0030] In some embodiments of the invention, said digital score is scrolled on the screen of the electronic device as a continuous string of musical signs, drawn on the screen in a consecutive way, by showing on the screen additional musical signs of music while the already scrolled musical signs disappear from the screen, in such a way that additional musical signs start to gradually appear on the screen while the already scrolled musical signs start to gradually disappear from the screen.

[0031] According to a second aspect of the present invention, there is provided a device comprising means for carrying out the method according to any preceding claim, said device being a personal digital assistant (PDA), a portable reader device, a tablet, a cell phone or any device which comprises a memory, a processor and a screen or display.

[0032] According to a third aspect of the present invention, there is provided a computer program product comprising computer program instructions/code, for performing the method already disclosed.

[0033] According to a fourth aspect of the present invention, there is provided a computer-readable memory/medium that stores program instructions/code, for performing the method already disclosed.

[0034] Additional advantages and features of the invention will become apparent from the detail description that follows and will be particularly pointed out in the appended claims.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0035] To complete the description and in order to provide for a better understanding of the invention, a set of drawings is provided. Said drawings form an integral part of the description and illustrate an embodiment of the invention, which should not be interpreted as restricting the scope of the invention, but just as an example of how the invention can be carried out. The drawings comprise the following figures:

Figure 1 shows on the left a representation of a typical audio signal produced by a guitar (up) and a typical audio signal produced by a clarinet (down). On the right, a graphic showing a combined audio signal produced by the simultaneous playing of a guitar and a clarinet.

Figures 2A to 2D show an example (four sequences) of vertical scrolling.

Figures 3A to 3E show an example (five sequences) of horizontal scrolling.

Figure 4 shows a virtual representation of the continuous scrolling of musical signs, in opposition to page-by-page scrolling.

Figure 5 shows a block diagram of the method for obtaining in real time the tempo at which a musician is playing a song according to an embodiment of the invention.

Figure 6 shows a block diagram of a first stage for start detection of the method represented in Figure 5.

Figure 7A shows a block diagram of the calculation of an onset function for each frame of audio signal, according to an embodiment of the invention. Figures 7B to 7D graphically represent the stage for onset detection of the method represented in Figure 7A.

Figures 8A to 8C graphically represent the stage for dominant period detection of the method represented in Figure 5.

Figure 9 shows a chart representing the tempo value (bpm) of audio frames of a musical performance. Each audio frame is represented as a vertical line. While usually the tempo value is theoretically constant as long as there are no changes in tempo value or in reference figure in the score, the real tempo at which the musician is playing is in the form of a non-constant continuous line. Spots depicted on each line representing audio frames represent the tempo value estimated by the algorithm of the invention for each audio frame prior to the tempo value normalization stage.

## DESCRIPTION OF A WAY OF CARRYING OUT THE INVENTION

**[0036]** The following description is not to be taken in a limiting sense but is given solely for the purpose of describing the broad principles of the invention. Next embodiments of the invention will be described by way of example, with reference to the above-mentioned drawings showing apparatuses and results according to the invention.

**[0037]** The method of displaying on the screen of an electronic device a score kept in a digital file is as follows: Preferably, a file having a digital score has been loaded in the memory of the device and the contents of the file stored in the buffer are read. Then, the total length of the score may be calculated in order to, by default, for example display the full score. Alternatively, only a first portion of the total length of the score may be calculated in order to, for example, display the first portion of the score. In this case, a second portion of the total length of the score is calculated before the first portion thereof has been played, and then displayed. This calculation and displaying of subsequent portions of score may be repeated until a last portion of the whole score is calculated and displayed. The width of the digital score may be adapted to that of the screen on which it is displayed. In other words, by default, as many music lines as required may be shown/drawn, in order to show on the screen all the notes of the score along the width of the screen. Since, however, for practical reasons, only a certain amount of "lines" can be shown on the screen -for the user to be able to read them-, a scrolling or displacing function is activated.

**[0038]** Once the contents of the file stored in the buffer are read, if the score has repetitions (portions to be played several times), the repetitions may be expanded. This means that those measures -or in general, musical signs- that should be played more than once, are concatenated in a row as many times as repetitions marked in the score, according to a specific notation in the score. The annotations corresponding to repetitions are marked in the digital file. Thanks to these marks, the algorithm, embedded in processing means, knows which portions must be expanded and how many times they must be expanded, that is to say, copied in a concatenated way. This process may fill the buffer with the score fully "expanded". In this process, a pre-buffer may be stored in a temporary buffer for subsequent use.

**[0039]** The user can therefore read and interpret music in a linear fashion, avoiding the need of going back in the digital score. An example of vertical scrolling is shown in figures 2A to 2D, wherein four sequences of a digital score being scrolled from bottom to top are illustrated. An example of horizontal scrolling is shown in figures 3A to 3E, wherein five sequences of a digital score being scrolled from right to left are illustrated. In vertical scrolling, musical symbols or signs move along an "y" axis (along the height of the screen), while in horizontal scrolling musical symbols or signs move along an "x" axis, (along the width of the screen). Figure 4 shows a virtual representation of the continuous scrolling according to embodiments of the invention, in which the notes or measures move along the screen (either from bottom to top or from right to left, that is to say, along a dimension of the screen).

**[0040]** The method of the present disclosure is then performed. Next it is explained how the scrolling speed is adjusted in real time to the music being played by the musician. The algorithm adapts the speed at which the digital score is shown, that is to say, scrolled on the screen of the electronic device, based on an estimated tempo at which the musician(s) is(are) playing the score. In other words, the algorithm is capable of estimating the tempo at which the musician(s) is(are) playing and of adjusting the scrolling speed of the digital score to the estimated tempo. From now on, the expression "the musician" generally refers to a single musician or to a group of musicians playing together. Thus, the musical signs scroll on the screen at the tempo at which the musician is playing. Following the tempo at which the musician is playing, the algorithm calculates the speed at which music should move on the screen, either vertically or horizontally, in such a way that the user is able to read it and interpret it, thus playing his/her instrument without interruptions and in a linear

way, as illustrated for example in Figure 4. In order to estimate in real-time the tempo at which the musician is playing, an algorithm for signal processing is applied, as explained next.

5 [0041] Once a first portion of digital score is drawn on the screen of the electronic device, a musician starts playing the song. The musician may then activate the algorithm for scrolling the digital score, for example by pressing a "start" button on the display prior to starting to play or by stepping on a pedal. Then, a method is performed for continuously obtaining or estimating a tempo at which the musician is playing, in order to adjust the scrolling speed to the estimated tempo. In this context, the term "continuously" refers to repeatedly recalculating the estimated tempo for frames of audio signal of certain time duration, as explained next. The music being played is sound waves. The music (sound waves) being played is captured by an audio sensing means, such as a microphone, embedded or connected to the electronic device on which screen the digital score is being displayed. The audio sensing means converts the captured sound into an analog audio signal. While it is captured, the analog audio signal is converted into a digital audio signal for example by means of an A/D converter. Some electronic devices may comprise processing means for producing a digital audio signal in a single step, transparent to the user. For example, the audio sensing means may be integrated or embedded together with analog-to-digital conversion means. From now on in this disclosure, the term "audio signal" refers to the already digitized analog signal.

10 [0042] Figure 5 shows a general block diagram of the method stages for obtaining an estimated tempo 504 at which the musician is playing and for adjusting the scrolling speed 53 taking into account the estimated tempo 504. Figure 5 represents a signal processing block for treating an audio signal 501 as captured by the audio sensing means and fully digitized, and corresponding to a musical performance. The method is executed in three stages: In a first stage 51, the rate at which music is being played is obtained; in other words, a dominant tempo value 502, also referred to as dominant rate, expressed in beats per time unit, typically BPM (beats per minute) is continuously obtained from the audio signal 501. In a second stage 52, an estimated tempo of the performance 504 is continuously obtained from the dominant tempo value 502 and from a reference tempo 503. In this stage 52, the dominant tempo value 502 is continuously normalized to a reference tempo 503, as a result of which the estimated tempo 504 of the music played by the musician is continuously obtained. These two stages 51, 52 are explained next. Finally, the scrolling speed is adjusted 53 from the estimated tempo 504. The method is repeated continuously until the end of the performance.

20 [0043] Figure 6 shows a block diagram of a stage 51 for obtaining or detecting a dominant tempo value 502 or dominant rate, such as dominant BPM, from an audio signal 501 corresponding to the music being played by the musician, according to a possible embodiment of the invention. In the context of the present disclosure, the term "dominant", referred to a tempo value, rate or BPM, means "the most repeated", such as the most repeated tempo value, obtained from the most repeated time interval between the strongest peaks in an onset function, in other words, from the most repeated periodicity, as will be explained later in this disclosure. First, a stage of onset function detection 511 is continuously applied to portions of the audio signal 501 being captured by the audio sensing means. This is done for the whole audio signal 501 (that is to say, until the end of the performance). Then, a stage of detection 512 of dominant tempo value, also referred to as dominant period detection, is performed. The onset detection 511 is applied in the form of a loop that lasts the duration of the score. In other words, for the whole song being played, frames frame<sub>i</sub> of the audio signal 501 of certain time duration are captured and then analyzed. The capture of frames is represented by reference 510 in figure 6. The time duration of the captured frames may be constant or non-constant. The time duration of these frames may be selected to be between 1 and 20 s (seconds), such as between 2 and 15 s, such as between 2 and 10 s. This selected time duration may be different for different users, electronic devices or other circumstances. For example, it may vary, depending on the processing resources of the electronic device, among other reasons. In embodiments in which the audio signal 501 is the analog signal obtained from the audio sensing means, the analog signal 501 is digitized, for example by means of an analog-to-digital converter (A/D converter) either prior to the capture of frames 510 or after such capture. In other embodiments, the audio sensing means (not shown) includes or is embedded together with, an A/D converter, as a consequence of which the audio signal 501 is already a digital signal.

30 [0044] Prior to starting to select frames of the audio signal 501 for the onset detection 511, the algorithm may check whether or not the musician has started to play the score displayed on the screen of the device. This verification may be done in different ways. For example, but not in a limiting way, it may be done by comparing energy levels, such as by comparing the mean energy of an audio signal frame with the mean energy of a reference audio signal frame or with the mean energy of a group of audio signal frames. If the result (difference) of this comparison is above a certain threshold, it may be determined that the musician has started to play. Alternatively, it may be done by knowing the frequency of the first note in the score and the tuning of the instrument being played. Other ways of verification may be used. The way this verification is performed is out of the scope of the present disclosure.

35 [0045] For each captured frame frame<sub>i</sub> of audio signal 501, an analysis of the captured frame frame<sub>i</sub> is performed in order to detect the onset 511 of each played note within said frame frame<sub>i</sub>. An onset occurs every time a musical note starts to play. An onset is represented as a peak in the temporal domain. Thus, for each frame frame<sub>i</sub> of audio signal 501, an onset function 61 is obtained, the onset function being a vector having the detected onsets and having the same temporal duration as the frame frame<sub>i</sub>. Because identifying in real time which specific note is being played

is practically impossible, as explained with reference to Figure 1, the method of the present disclosure analyses the audio signal frame by frame in order to identify in each frame the tempo value, that is to say, the rate at which the musician is playing. In particular, as shown in Figure 7A, a spectrogram is computed 71 from each frame  $frame\_i$  into which the audio signal 501 is divided. The spectrogram represents the spectrum of frequencies in the audio signal -or rather, in a frame  $frame\_i$  of the audio signal-. It represents how the frequencies vary with time (frequency on the vertical axis, time on the horizontal axis). The spectrogram may be obtained by means of Fourier transform. The spectrogram may be obtained for example, but not limiting, using the FFT (Fast Fourier Transform). The spectrogram represents a time-frequency energy distribution of the audio signal in a given analysis frame. Analytically, it can be obtained as a squared modulus of short-time Fourier transform (STFT) of the audio signal in a given analysis frame.

**[0046]** In Figure 7B a frame  $frame\_i$  of the audio signal is shown, from which a spectrogram is calculated in block 71. In figure 7C a spectrogram is illustrated. The spectrogram has been obtained by Fourier transform, such as a Fast Fourier transform (FFT), of the frame shown in figure 7B. As a skilled person is aware of, if the frames  $frame\_i$  correspond to an analog audio signal, the time-domain frames are sampled for digital conversion and then Fourier transform is performed on each group of samples. In Figure 7C, a third dimension indicating the amplitude of a particular frequency at a particular time is represented by the intensity of color (in grey scale) of each point in the image. Then, for each frame  $frame\_i$ , and therefore for each spectrogram obtained therefrom, an onset function 61 is obtained (block 72 in figure 7A). At block 72, each spectrogram is processed as follows: First, a vector of weights is generated in order to recombine the samples of the spectrogram into Mel-frequency bands. For example, the samples may be recombined into 40-channel Mel-frequency bands. The result of this operation is called Mel-spectrogram, as disclosed for example by Haytham Fayek in Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between, April 21, 2016 (<http://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>). Next, the Mel-spectrogram is logarithmically compressed. In particular, the Mel-spectrogram samples may be calculated in decibels and those smaller than a certain threshold are rejected (set to zero). The threshold may be fixed, for example, to -80 decibels. The result of this operation is called Mel-log-spectrogram. Finally, the discrete time difference is compute don the Mel-log-spectrogram and only the positive values are retained, because they correspond to note onsets. These note onsets form the onset function 61. In Figure 7D, the peaks in the temporal domain, which are the onset function 61, look like the changes of note in the score on which the audio signal is based.

**[0047]** Although it has been observed that in many situations the results 61 of the onset detection stage 511 are reliable, in the sense that the peaks in the temporal domain (figure 7D) correctly correspond to the onset function of the played portion of music, there is a relatively high error probability, due to several reasons, such as low signal-to-noise ratio (SNR) for example when the musician plays quietly, or low temporal resolution (if there is only a few bits in the window duration (frame duration), or if there are long rests, or for any other reason. In order to minimize this error probability, a stage 512 for detection of dominant tempo value is applied to each onset function 61 (that is to say, after the onset detection stage 511 for each captured frame  $frame\_i$  of audio signal). The stage 512 for dominant tempo value detection is preferably implemented as an autocorrelation function. This is done for maximizing the probability of finding a dominant period in each sample  $frame\_i$ , that is to say, of finding the most repeated time interval between the strongest peaks in the onset function 61. From the time interval or period T between the strongest peaks in the onset function, the dominant tempo value (BPM) -also referred to as dominant BPM- of the captured frame  $frame\_i$  is directly obtained as follows: dominant tempo value =  $1/T$  (minutes) =  $60/T$  (seconds).

**[0048]** Figure 8C graphically represents the autocorrelation signal 55 obtained after applying an autocorrelation function (stage 512 for dominant period detection in figure 6) to the onset function 61 shown in figure 8B in turn obtained from a frame  $frame\_i$  of audio signal 501 shown in figure 8A. In this particular example, figures 8A and 8B correspond to a frame of 5 seconds duration. The autocorrelation function may be a discrete time autocorrelation function. The strongest peak in the autocorrelation signal represents the dominant periodicity and therefore the dominant tempo value 502 of the audio signal analyzed in each frame. In the illustrated figures, as a matter of example, the strongest peak (period T) occurs at 0.35 s (=0.00583 minutes). Therefore, the tempo value of the analyzed signal frame is  $1/0.00583 = 171$  BPM.

**[0049]** Turning back to figure 5, so far, a dominant tempo value (current tempo value of the performance as captured in  $frame\_i$ ) 502 is obtained for every frame  $frame\_i$  of audio signal 501. In order to match this dominant tempo value with the content of the digital score being played, and therefore to adjust the speed at which the digital score scrolls, for each frame  $frame\_i$ , the dominant tempo value 502 must be normalized with respect to a reference tempo 503. In other words, for each frame  $frame\_i$ , an estimated tempo 504 of the performance as represented by  $frame\_i$  must be obtained from the dominant tempo value 502 and from a reference tempo 503. That is to say, the actual tempo of the music being played must be obtained. In particular, the reference musical figure or simply reference figure of the reference tempo 503 is required. Theoretically, as long as there are no changes in tempo value or in reference figure in the score, the evolution of the tempo in a musical piece is constant (the tempo should not vary). However, tempo is in fact a continuous function of slow and smooth variation (without remarkable steps). Figure 9 shows the tempo value (bpm) of audio frames of a musical performance. In the "x" axis audio frames  $frame\_i$  are represented (1, 2, 3, 4...) as vertical lines. In the "y" axis the tempo value (bpm) is represented. In dotted line, a theoretical constant tempo value of a portion of the musical

performance is represented. In continuous line, the actual tempo value, at which the musician is playing, which is not constant, is represented. For each audio frame, a spot represents the dominant tempo value 502 calculated prior to the dominant tempo value normalization stage 52 by the method of the present disclosure. In this example, the errors in the estimation of frames 1 and 3 are most likely errors caused by the tempo estimation itself (because a musician cannot follow a determined tempo value with absolute precision), while the errors in the estimation of frames 2 and 4 are most likely errors caused by the rhythmic ambiguity of the frame. In order to compensate for these errors (frames 2 and 4), the dominant period 502 must be normalized to a reference figure.

**[0050]** Next, it is explained how, for each frame  $frame\_i$  of the audio signal 501, the calculated dominant tempo value 502 is normalized to a reference tempo 503, in order to update the estimated tempo 504 of the performance.

**[0051]** At the beginning of the performance, that is to say, for example when the algorithm for scrolling the digital score is activated, a reference tempo (tempo value & reference figure) is often indicated or suggested. For example, a reference tempo may be suggested on the score, typically using an Italian word (*Andante...*) associated to certain predetermined or well-known value, or a reference tempo is imposed by the conductor. In other occasions, the player chooses the reference tempo at which he/she is going to play. Next, non-limiting examples of ways of establishing a reference tempo are described. The reference tempo may be manually fixed by the musician, for example by typing it on the screen of the electronic device in order for the algorithm to be aware of the reference tempo. This is an option available at the APP implementing the algorithm. In this case, a reference figure extracted from a time signature in the score may be taken into account, or alternatively the user may freely establish the reference tempo in BPM at his/her will. The reference tempo may be extracted (obtained) from the digital score, for example as a metronome mark (for example quarter note = 50). The reference tempo may also be obtained by combining the two former possibilities, that is to say, using the Italian word indicated on the score together with a metronome mark or as time signature, or as indication of metric changes along the score, among other ways of obtaining the reference tempo from the digital score. The reference tempo may also be provided by the APP by default. The reference tempo may also be extracted from a data base. Regarding the time signatures, as a skilled person is aware of, the reference tempo may be extracted therefrom in different ways, depending on the types of measures: In binary measures, such as 2/2, 2/4, 4/4 and so on, the reference figure is determined by the denominator of the fraction (2 = half note, 4 = quarter note, and so on). In ternary measures, such as 3/8, 6/8, 9/8, 12/8 and so on, groups of 3 figures are made, the denominator indicating the type of figure. For this reason, in order to know the amount of beats of a ternary measure, the numerator must be divided by 3. For example, in a 3/8 measure there is 1 beat, in a 9/8 measure there are 3 beats. The reference figure is determined by the denominator. In these examples (3/8, 6/8, 9/8, 12/8...) the denominator indicates "eighth note". Therefore, because 3 eighth notes must be grouped in each beat, the reference figure is the sum of 3 eighth notes, that is to say, a dotted quarter note.

**[0052]** Thus, the reference tempo indicated by any of the above mentioned ways, or by any other way, is the reference tempo 503 for the first audio signal frame  $frame\_1$ . So, when the musician starts to play, music is captured by the audio sensing means, such as a microphone, as audio signal. From the first captured frame  $frame\_1$  of the digitized audio signal 501, a dominant tempo value 502 is obtained. Then, from this dominant tempo value 502 and from the reference tempo 503, the estimated tempo 504 of the first portion of the performance (corresponding to the first frame  $frame\_1$ ) is estimated.

**[0053]** For the subsequent frames ( $frame\_2$ ,  $frame\_3$ ,  $frame\_4$ ... in general  $frame\_i$ ) of the audio signal 501, the reference tempo 503 may be the estimated tempo of the previous frame  $frame\_i-1$ . Or the reference tempo 503 may be an average estimated tempo calculated taking into account a certain number N of previous frames. Or the reference tempo 503 may be indicated to the algorithm by any of the ways already enumerated.

**[0054]** Next, it is explained how the estimated tempo 504 of the performance may be obtained, frame by frame, in embodiments of the invention.

**[0055]** In order to compensate for errors derived from rhythmic ambiguity, a set of scaling values are applied to the dominant tempo value 502 in order to normalize the dominant tempo value 502 to the reference tempo 503. In a particular

embodiment, the set of scaling values are: 
$$p = \left\{ 3, 2, \frac{3}{2}, 1, \frac{1}{2}, \frac{1}{3} \right\}.$$
 These values are related to the possible musical figures that most probably may dominate in an audio frame. The mentioned set of scaling values imply that corresponding musical figures may take the estimated tempo value to the triple, double, 3/2 times, same, half, or one third of the actual tempo value. In another particular embodiment, the set of scaling values are:  $p = \{4, 3, 2, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}\}.$

**[0056]** So, for each dominant tempo value 502 (there is one dominant tempo value 502 per audio frame  $frame\_i$ ), the dominant tempo value 502 is multiplied by all the values of the set of scaling values, thus obtaining a set of scaled dominant tempo values. Next, the absolute difference between each value of the set of scaled dominant tempo values and the reference figure of the reference tempo 503 is calculated:  $|\text{reference figure} - p * \text{dominant BPM}|.$  The modulus (positive value) of the difference is considered. The lowest absolute difference indicates the scaling value by which the dominant tempo value 502 must be multiplied in order to obtain the tempo value of the estimated tempo 504, the reference figure of the estimated tempo 504 being the reference figure of the reference tempo 503. When the dominant tempo

value 502 is multiplied by the selected scaling value, the dominant tempo value 502 becomes normalized to the reference figure 503. In other words, the tempo value of the estimated tempo 504 is the scaled dominant tempo value (scaled by the selected scaling value). This way, it is established that the tempo value of the estimated tempo 504 is the correctly scaled dominant tempo value and the reference figure of the estimated tempo 504 is the reference figure of the reference tempo 503.

**[0057]** For the second and subsequent frames (frame\_2, frame\_3, etc.), the reference tempo 503 may be based on the previously normalized tempos 504. For example, the reference tempo 503 for frame\_i may be the estimated tempo for frame\_i-1, or an average tempo calculated taking into account the last N frames (frame\_i-1, frame\_i-2, ...frame\_i-N). Alternatively, it may be decided that the reference tempo 503 for frames other than the first one frame\_1 is based on an indicated reference tempo, for example indicated to the algorithm by any of the ways already enumerated.

**[0058]** Next a first example of tempo estimation 504 is disclosed, in which the set of scaling values is  $\left\{3, 2, \frac{3}{2}, 1, \frac{1}{2}, \frac{1}{3}\right\}$ . The musician has just started playing and therefore the first frame frame\_1 of the audio signal 501 has just been captured. A reference tempo 503 has been provided by the musician (for example by typing it on a window opened with that purpose on the screen of the device). The reference tempo is quarter note = 120. For frame frame\_1, the value of the dominant tempo value 502 has been computed and is 60. This value 60 of dominant tempo value 502 is multiplied by

all the values of the set of scaling values  $p = \left\{3, 2, \frac{3}{2}, 1, \frac{1}{2}, \frac{1}{3}\right\}$ , thus obtaining the following set of scaled dominant

tempo values:  $\left\{3 * 60, 2 * 60, \frac{3}{2} * 60, 1 * 60, \frac{1}{2} * 60, \frac{1}{3} * 60\right\} = \{180, 120, 90, 60, 30, 20\}$ . Next, the absolute difference between each value of the set of scaled dominant tempo values and the reference tempo (quarter note = 120) is calculated: {60, 0, 30, 60, 90, 100}. The lowest absolute difference is in this case 0, corresponding to the second scaling value. Therefore, the dominant tempo value 502 (60 in this example) must be multiplied by the scaling value 2. Thus, the normalized tempo value of the estimated tempo 504 for this frame frame\_1 is  $2 * 60 = 120$ . And the reference figure of the estimated tempo 504 is the reference figure of the reference tempo 503, that is to say, "quarter note". The dominant tempo value 502 is thus normalized to the reference figure 503.

**[0059]** Next a second example of tempo estimation 504 is disclosed. The musician keeps on playing and a second frame frame\_2 of the audio signal 501 has just been captured. The reference tempo 503 is now the estimated tempo 504 of the previous iteration (frame\_1) which, according to the first example, is quarter note = 120. For frame frame\_2, the value of the dominant tempo value 502 is 123. This value 123 of dominant tempo value 502 is multiplied by all the

values of the set of scaling values  $p = \left\{3, 2, \frac{3}{2}, 1, \frac{1}{2}, \frac{1}{3}\right\}$ , thus obtaining the following set of scaled dominant BPMs:

$$\left\{3 * 123, 2 * 123, \frac{3}{2} * 123, 1 * 123, \frac{1}{2} * 123, \frac{1}{3} * 123\right\} = \{369, 246, 184.5, 123, 61.5, 41\}.$$

Next, the absolute difference between each value of the set of scaled dominant tempo values and the tempo value (in this case, 120) of the reference tempo 503 is calculated: {249, 126, 64.5, 3, 58.5, 79}. The lowest absolute difference is in this case 3, corresponding to the fourth scaling value. Therefore, the dominant tempo value 502 (123 in this example) must be multiplied by the scaling value 1. Thus, the normalized tempo value of the estimated tempo 504 for this frame frame\_2 is  $1 * 123 = 123$ . Thus, the normalized tempo value of the estimated tempo 504 for the second frame frame\_2 is  $1 * 123 = 123$ . And the reference figure of the estimated tempo 504 is the reference figure of the reference tempo 503, that is to say, quarter note. The digital score is scrolled on the screen of the electronic device at a speed adjusted 53 from the estimated tempo 504 at which the player is actually playing. The scrolling speed is recalculated -adjusted- every time a new frame frame\_i of the audio signal is captured. In other words, the scrolling speed is recalculated in real time, since a new frame is captured and analyzed every few seconds or even milliseconds.

**[0060]** In some cases, it may happen that there is not a clear minimum value in the set of absolute differences. The reason for this may be that the dominant tempo value 502 may have been erroneously calculated. In this case, in embodiments of the invention, the algorithm reacts to this mistake and discards the estimation performed for the current frame (for example frame\_j). The estimated tempo for the current frame frame\_j is replaced, for example, for a mean value of all the previous estimated tempos (from frame\_1 to frame\_j-1), or for a mean value of the N previously estimated tempos (from frame\_j-N-1 to frame\_j-1) or for the lastly estimated tempo (for frame\_j-1). In order to detect this kind of events, after the absolute difference between each value of the set of scaled dominant tempo values and the tempo value of the reference tempo 503 has been calculated, the two lowest values of the set of absolute differences are selected. These two values indicate that the audio frame under analysis may probably comprise musical figures having as rhythmical value the musical figures (tempo values) represented by the two positions associated to those two selected

values in the set of scaling values, for example  $\left\{3, 2, \frac{3}{2}, 1, \frac{1}{2}, \frac{1}{3}\right\}$  or  $\{4, 3, 2, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}\}$ , considering a reference tempo. In other words, two musical figures are selected as candidates (for example the musical figures represented by "2" and "3/2"). A ratio R that limits the deviation of the current estimated tempo from an average value of previous estimated tempos is then used. This ratio R is applied as a threshold, as follows: the two lowest values already selected are divided  $V_1/V_2$ , wherein  $V_1$  is the lowest value of the two values and  $V_2$  is the highest value of the two values. If  $V_1/V_2 \leq R$  (ratio previously established), then the estimation is considered correct and it is considered that the audio frame has a rhythmical value inverse with respect to the musical figure corresponding to the position of the lowest value  $V_1$  in

the set of scaling values, such as, for example,  $\left\{3, 2, \frac{3}{2}, 1, \frac{1}{2}, \frac{1}{3}\right\}$  or  $\{4, 3, 2, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}\}$ . So, for that frame, the adjusted tempo value of the estimated tempo is calculated by multiplying the estimated tempo value by the scaling value corresponding to that position. If, on the contrary,  $V_1/V_2 > R$ , then it is considered that the estimated tempo value has deviated too much from any potential tempo as a consequence of a severe error. The estimated tempo value is then considered to be a different one, for example a mean value of all the previous estimated tempo values (from frame\_1 to frame\_j-1), of a mean value of the N previously estimated tempo values (from frame\_1+N to frame\_j-1) or the lastly estimated tempo value (for frame\_j-1). The rate R may be empirically obtained. For example, the rate R may be selected to be  $0.6 < R < 0.9$ . The rate R may be fixed, for example, but not limiting, to  $R = 0.8$ .

**[0061]** In sum, if  $V_1/V_2 \leq R$ , it is considered that there is no error and the tempo value of the estimated tempo 504 is calculated following the general method. If, on the contrary,  $V_1/V_2 > R$ , the correction disclosed in this paragraph is applied.

**[0062]** So, for each audio frame frame\_i, an adjusted estimated tempo 504 is calculated. And the digital score is scrolled on the screen of the electronic device at a speed adjusted 53 from the adjusted estimated tempo 504 at which the player is actually playing. The scrolling speed is recalculated every time a new frame frame\_i of the audio signal is captured. The scrolling speed is adjusted as follows: The scrolling speed is adjusted according to the musical signs being displayed on the screen at each time instant (for example, every time a new frame frame\_i is captured) and to the obtained estimated tempo 504. For each displayed sign, the time length or time duration required for playing the displayed sign is calculated as the amount of reference musical figures in the sign divided by the tempo value. In other words, the time length or time duration needed by the sign to cover the length or width (depending on whether the scroll is vertical or horizontal) of the screen is calculated as the amount of reference musical figures in the sign divided by the tempo value. If required, the dimensions (length or width) of the screen (space to be covered by a musical sign) may be obtained from the electronic device.

**[0063]** As apparent from the description above, the invention provides many advantages over prior art methods of scrolling musical scores. Some advantages are recited next: It is not necessary to establish the exact point of reading, but only the beats per minute at which the player is playing. Therefore, different rhythmic variations are detected without causing a reaction critical point. Thus, the scrolling speed variations necessary to adapt the displayed portion of score to the actual tempo of the player have a wide margin and can be soft, without compromising at any moment the reading of the score. The flow of the digital score on the screen itself acts as a "time line", but it leaves at all times the player a wide margin for reading both ahead and behind the exact point of interpretation. The tempo of any performance can be detected, even from a performance obtained from a recording or from a live concert. This is because no tuning or detection of specific music tones is required. Therefore, it is able to detect and recalculate at all times the tempo value or "BPM" (beats per minute) of any musical score or performance by setting a reference musical figure. Therefore, because the number of beats shown on the screen at every moment is known, the score can be scrolled on the screen according to the real time tempo of the actual performance or recording. For the same reason, and because all the musicians belonging to a group (for example an orchestra) should follow a same rhythm of music interpretation, detecting the tempo value or "BPM" of a single musician, without taking into account tunings or particular tones of different instruments, enables the detection of the tempo of a group of instruments, each of which do not cause interference in the sound detection of the others. Therefore, the method can be carried out by a plurality of users playing simultaneously the same score, but different particellas. In that case, each user has a device of the ones already described (at least with a processor, memory and screen), the digital score being shown in a device of each user. In this case, one of the devices can work as a master one, in the sense that the other devices synchronize with respect to this one. The remaining devices, however, keep the possibility of scaling the screen according to their needs (for example, visual needs). And because different electronic devices may be synchronized, all the particellas of different instruments forming for example an orchestra (or any other musical group) may be synchronized, in such a way that the scrolling of the score on each of them is done at the same speed, which is continuously recalculated from the detected tempo of the music being played. Alternatively, each electronic device may scroll the corresponding score (particella) independently -that is to say, not synchronized- from the scroll of the other devices in the musical group.

**[0064]** Concerning the scores, they can be stored either in the device itself (locally) or in an external site in the Internet

(cloud). In this last case, the user normally accesses this restricted area via identity name and password.

**[0065]** The software application also permits the user to purchase scores. Preferably, once a score has been purchased, it is stored in an external system restricted to a particular classification of metadata.

**[0066]** In this text, the term "comprises" and its derivations (such as "comprising", etc.) should not be understood in an excluding sense, that is, these terms should not be interpreted as excluding the possibility that what is described and defined may include further elements, steps, etc.

**[0067]** In the context of the present invention, the term "approximately" and terms of its family (such as "approximate", etc.) should be understood as indicating values very near to those which accompany the aforementioned term. That is to say, a deviation within reasonable limits from an exact value should be accepted, because a skilled person in the art will understand that such a deviation from the values indicated is inevitable due to measurement inaccuracies, etc. The same applies to the terms "about" and "around" and "substantially".

**[0068]** On the other hand, the invention is obviously not limited to the specific embodiment(s) described herein, but also encompasses any variations that may be considered by any person skilled in the art (for example, as regards the choice of materials, dimensions, components, configuration, etc.), within the general scope of the invention as defined in the claims.

## Claims

1. A method for displaying a musical score on a screen of an electronic device, comprising:

loading a file having a digital score in a piece of memory of said electronic device;  
 scrolling said digital score on the screen of the electronic device;  
 capturing an audio signal (501) corresponding to said musical score being played by a musician;  
 repeatedly selecting frames (frame<sub>i</sub>) of the captured audio signal (501) and, for each selected frame (frame<sub>i</sub>):

obtaining a dominant tempo value (502) at which the music contained in said frame (frame<sub>i</sub>) is played;  
 from the dominant tempo value (502) obtained from said frame (frame<sub>i</sub>) and from a reference tempo (503) comprising a reference figure and a reference tempo value, estimating (52) the tempo at which the musician is playing, said estimated tempo (504) comprising said reference figure and a normalized tempo value with respect to the dominant tempo value (502);  
 adjusting the scrolling speed of the digital score according to the estimated tempo (504).

2. The method of claim 1, wherein said dominant tempo value (502) at which the music contained in said frame (frame<sub>i</sub>) is played, is obtained as follows:

detecting (511) an onset function (61) of said frame (frame<sub>i</sub>);  
 finding (512) a dominant tempo value (502) in said onset function (61).

3. The method of claim 2, wherein the onset function (61) is detected (511) as follows:

obtaining (71) a spectrogram from the captured frame (frame<sub>i</sub>);  
 obtaining (72) the onset function (61) from the spectrogram.

4. The method of either claim 2 or 3, wherein the dominant tempo value (502) is obtained (512) by applying an autocorrelation function to the onset function (61).

5. The method of any preceding claim, wherein for a frame (frame<sub>i</sub>), the estimated tempo (504) obtained from the dominant tempo value (502) and from a reference tempo (503) is obtained as follows:

applying a set of scaling values to the obtained dominant tempo value (502), thus obtaining a set of scaled dominant tempo values,  
 calculating the absolute difference between each value of the set of scaled dominant tempo values and the reference tempo value of the reference tempo (503),  
 selecting the scaling value of the set of scaling values corresponding to the lowest absolute difference,  
 multiplying the dominant tempo value (502) by the selected scaling value, thus obtaining a normalized tempo value of the estimated tempo (504), the estimated tempo (504) comprising said normalized tempo value and the reference figure of the reference tempo (503).

6. The method of claim 5, further comprising:

after calculating the absolute difference, selecting the two lowest values  $V_1$ ,  $V_2$ , wherein  $V_1$  is the lowest value of the two values and  $V_2$  is the highest value of the two values;

if  $V_1/V_2 \leq R$ , wherein  $R$  is a ratio established for limiting the deviation of the current estimated tempo, the estimation is considered correct,

if  $V_1/V_2 > R$ , then the estimation is considered incorrect.

7. The method of claim 6, wherein if the estimation for the current frame (frame\_i) is considered incorrect, the estimated tempo for that frame is obtained from a number of previously obtained estimated tempos for corresponding previous frames.

8. The method of any preceding claim, wherein the reference tempo (503) used for obtaining (52) the estimated tempo (504) is obtained as follows: The reference tempo (503) is manually fixed by the musician, or the reference tempo (503) is obtained from the digital score, or the reference tempo (503) is provided by the algorithm by default, or the reference tempo (503) is extracted from a data base.

9. The method of claim 8, wherein the reference tempo (503) is obtained from the digital score as follows: as a metronome mark, as a value associated to a word included in the score, as a time signature, as indication of metric changes along the score, or as a combination of the former.

10. The method of any preceding claim, wherein the scrolling speed of said musical score on the screen is readjusted every time a new frame (frame\_i) of the audio signal is captured.

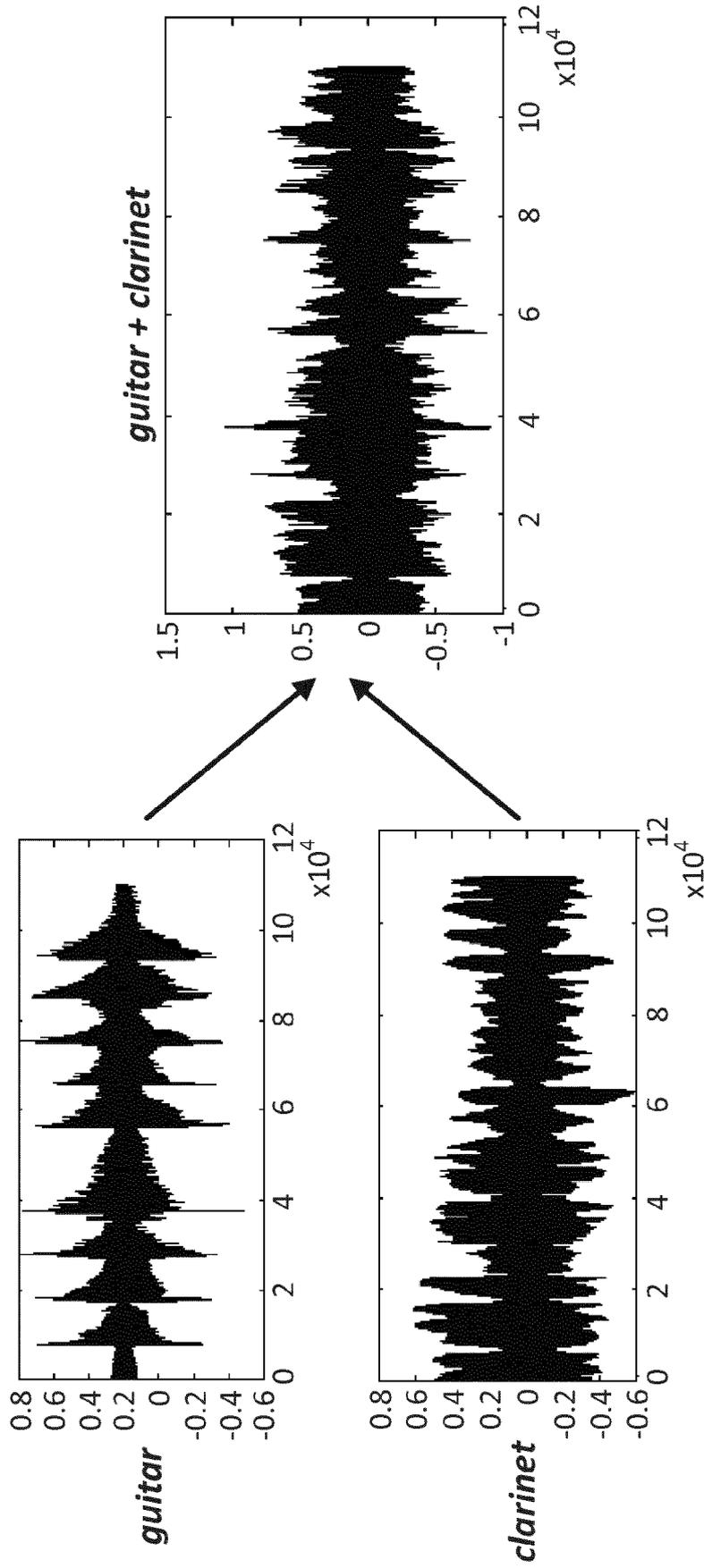
11. The method of any preceding claims, further comprising, prior to continuously capturing frames (frame\_i) of said audio signal (501), verifying that the first played notes correspond to a starting point identified in the digital score.

12. The method of any preceding claim, wherein said digital score is scrolled on the screen of the electronic device as a continuous string of musical signs, drawn on the screen in a consecutive way, by showing on the screen additional musical signs of music while the already scrolled musical signs disappear from the screen, in such a way that additional musical signs start to gradually appear on the screen while the already scrolled musical signs start to gradually disappear from the screen.

13. A device comprising means for carrying out the method according to any preceding claim, said device being a personal digital assistant (PDA), a portable reader device, a tablet, a cell phone or any device which comprises a memory, a processor and a screen or display.

14. A computer program product comprising computer program instructions/code, for performing the method according to any of claims 1-13.

15. A computer-readable memory/medium that stores program instructions/code, for performing the method according to any of claims 1-13.



**FIG. 1**

The image displays a musical score for a bass instrument, likely a double bass, in 4/4 time. The tempo is marked as "Largo". The score is divided into three measures:

- Measure 5:** Features a bass clef, a key signature of two flats (B-flat and E-flat), and a 4/4 time signature. The notation includes a thick horizontal bar across the staff, a measure rest, and a "5" above the staff. The tempo marking "Largo" is positioned above the staff.
- Measure 8:** Continues the bass clef and key signature. It features a measure rest, a half note, and a slur over a half note. A "7" is written below the staff.
- Measure 11:** Continues the bass clef and key signature. It features a measure rest, a half note, and a slur over a half note. A "7" is written below the staff.

At the end of the score, there is a tempo indicator showing a quarter note followed by "= 120" and a play button icon (a triangle inside a circle).

FIG. 2A

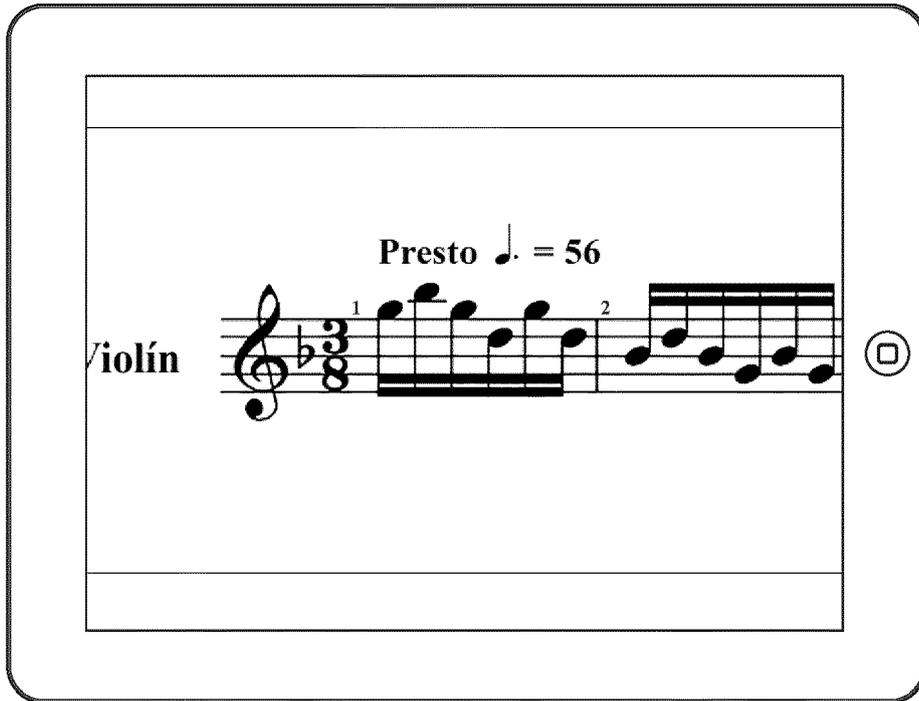
The image displays a musical score for FIG. 2B, consisting of three staves of music. The first staff begins with a bass clef, a key signature of three flats (B-flat, E-flat, A-flat), and a 4/4 time signature. The tempo is marked 'Largo'. A thick black bar covers the first five measures, with a large '5' centered above it. The second staff starts at measure 8, with a bass clef and the same key signature. It features a melodic line with a slur over measures 8 and 9, and a fermata over measure 10. The third staff starts at measure 11, with a bass clef and the same key signature. It contains a melodic line with a slur over measures 11 and 12. A tempo change box is located above the third staff, containing a quarter note followed by '= 120' and a play button icon. The entire score is set against a background of horizontal lines.

FIG. 2B

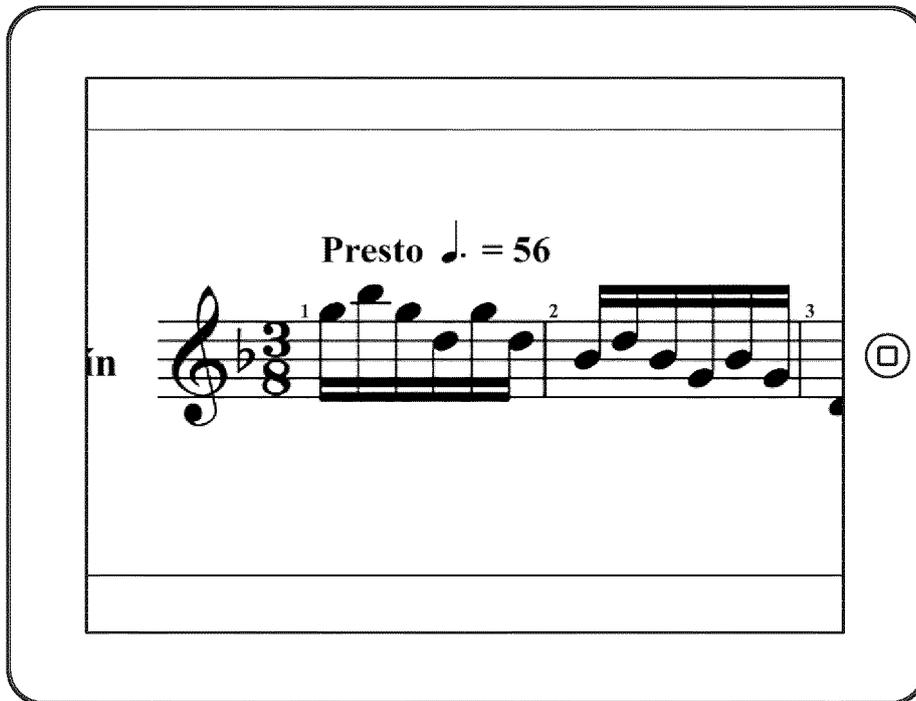


The image displays a musical score for FIG. 2D, consisting of three staves of music. The first staff begins with a bass clef, a key signature of three flats (B-flat, E-flat, A-flat), and a 4/4 time signature. A large number '5' is positioned above the staff. The tempo is marked 'Largo'. The second staff starts with a measure number '8' and features a slur over a group of notes. The third staff starts with a measure number '11' and includes a tempo marking '♩ = 120' with a right-pointing triangle symbol. The score concludes with a thick black bar at the bottom.

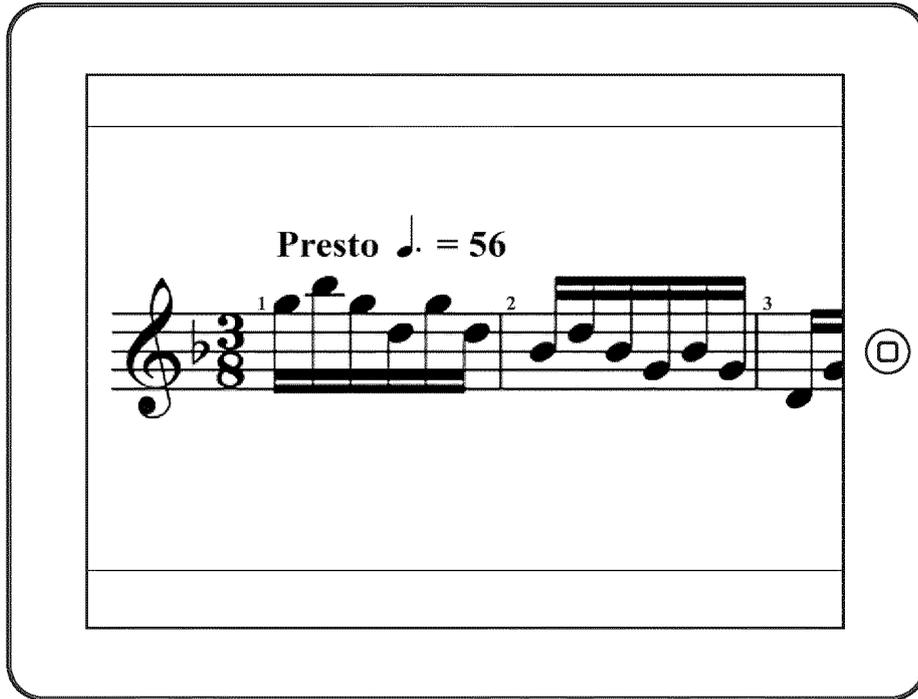
FIG. 2D



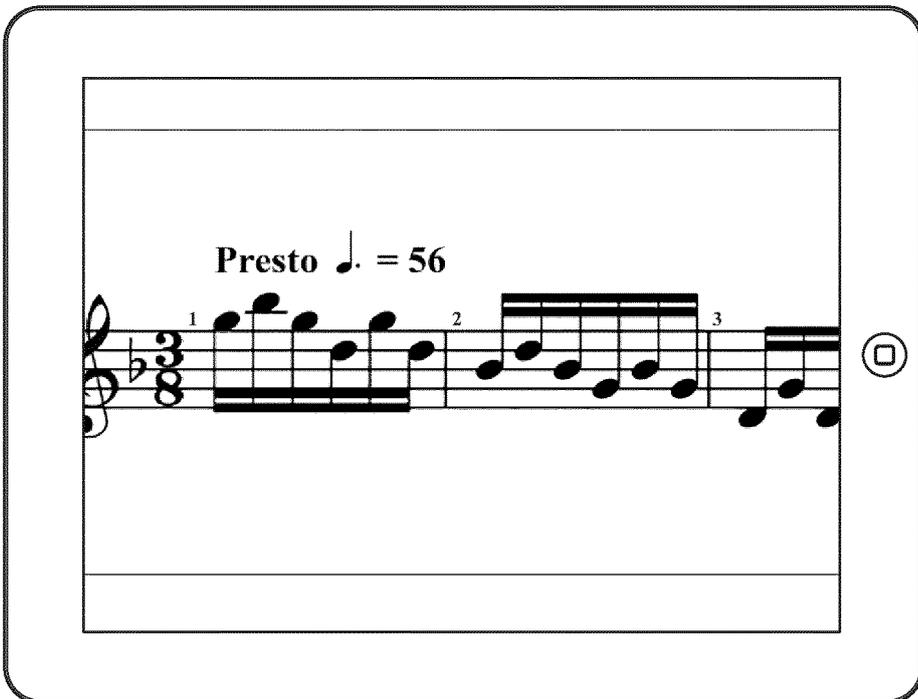
**FIG. 3A**



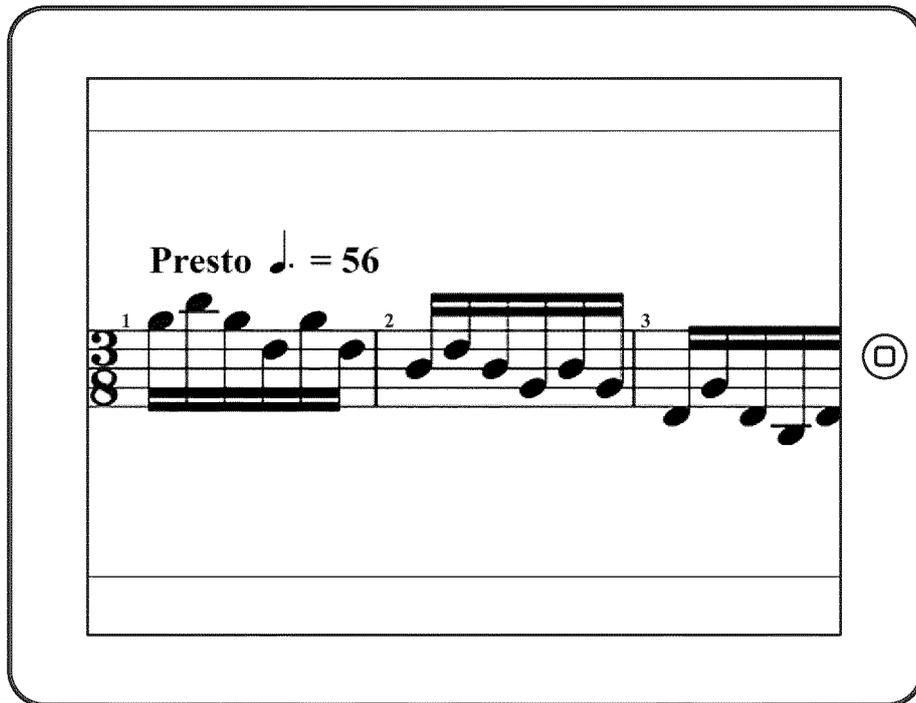
**FIG. 3B**



**FIG. 3C**



**FIG. 3D**



**FIG. 3E**

Em D C D

2 3 4 5

6 7 8 9 10

D Em G D G D

11 12 13 14 15

C D G D Em

16 17 18 19 20

G D G D C D

21 22 23 24 25

Em D C D Em

26 27 28 29 30

D C Bm Am Em D

36 37 38 39 40

Bm Am G

41 42

FIG. 4

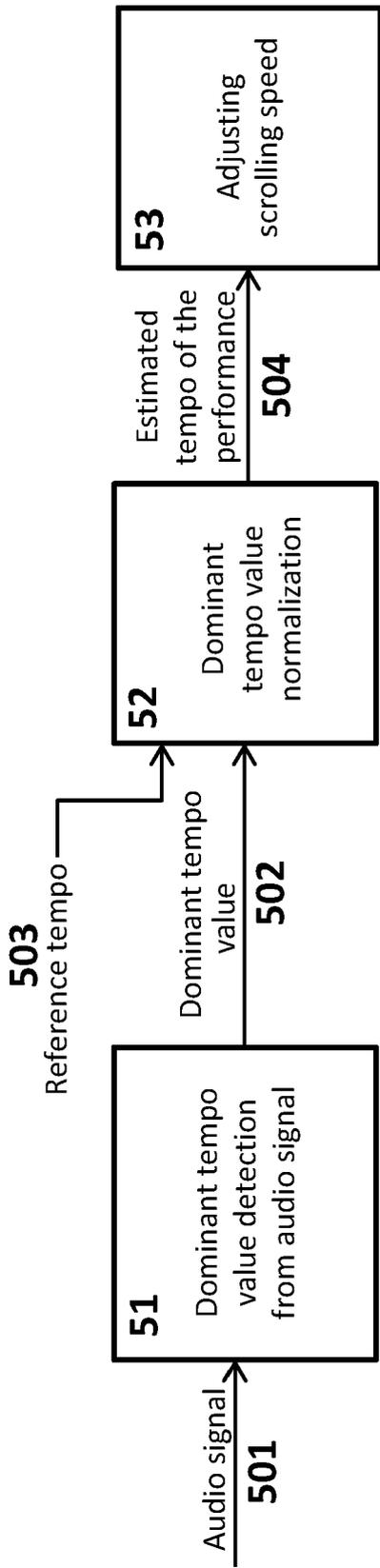


FIG. 5

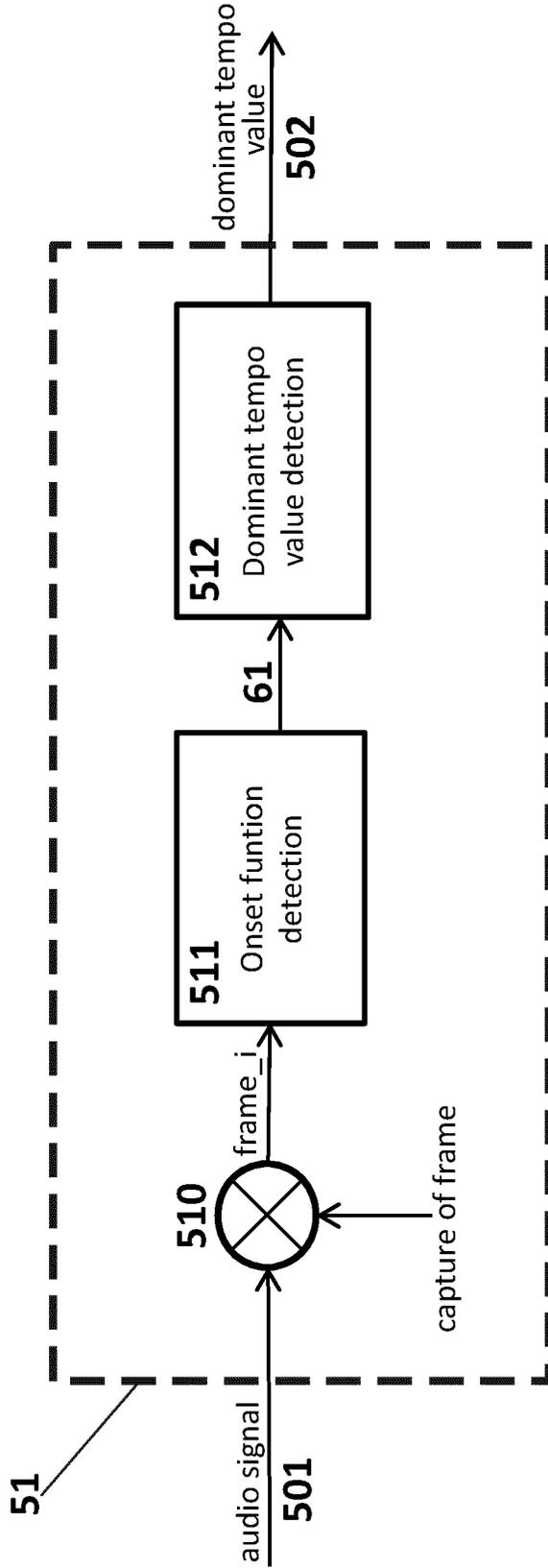


FIG. 6

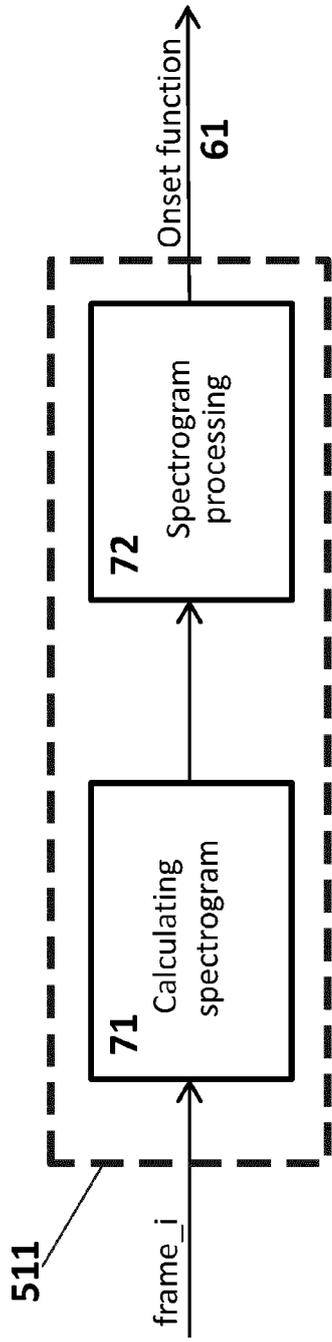


FIG. 7A

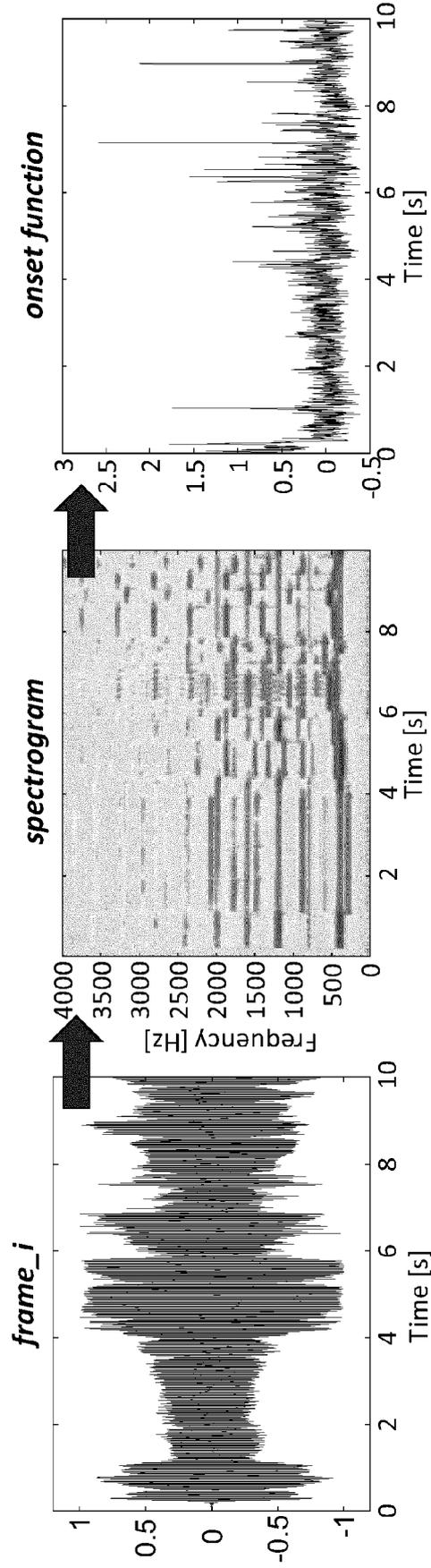
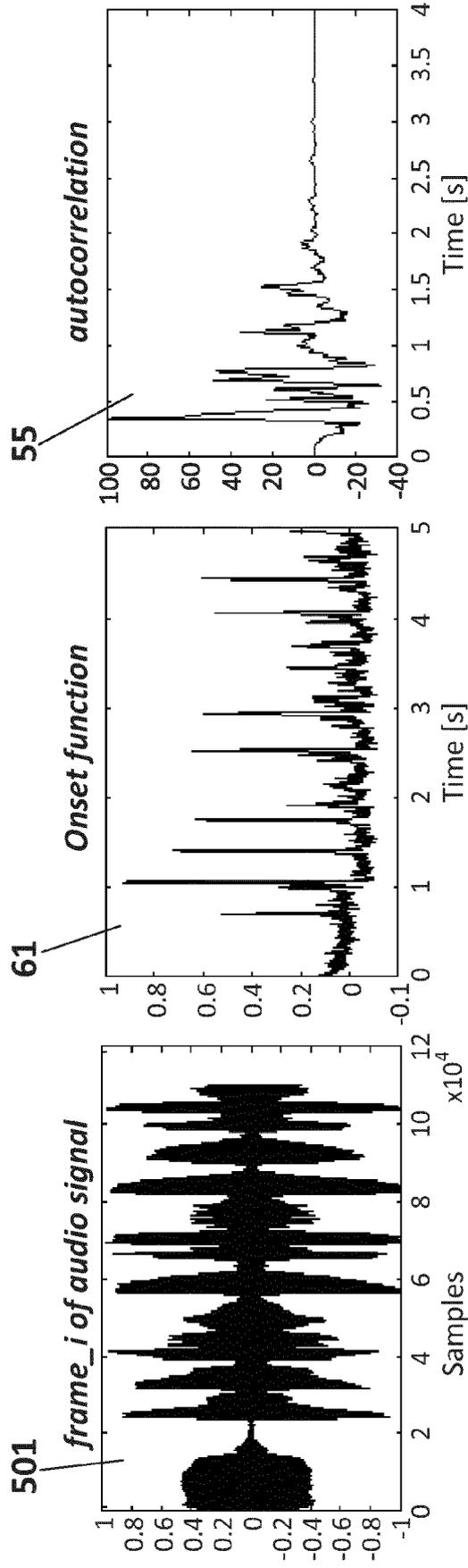


FIG. 7B

FIG. 7C

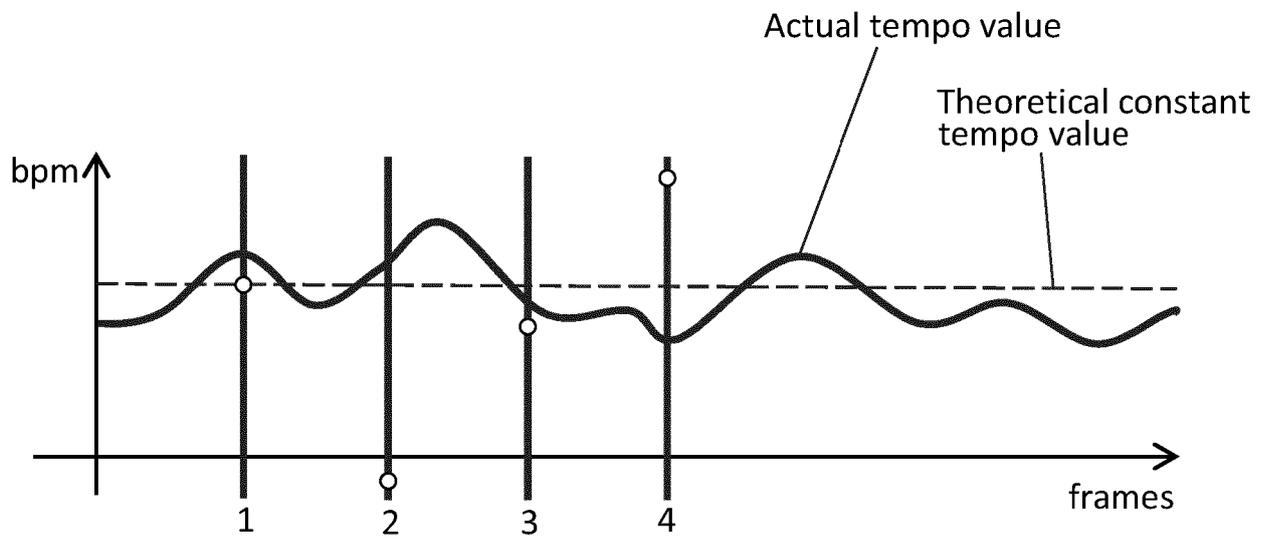
FIG. 7D



**FIG. 8C**

**FIG. 8B**

**FIG. 8A**



**FIG. 9**



EUROPEAN SEARCH REPORT

Application Number  
EP 18 38 2392

5

10

15

20

25

30

35

40

45

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	US 2017/110102 A1 (COLAFRANCESCO JULIEN [FR] ET AL) 20 April 2017 (2017-04-20) * abstract; figures 1-3 * * paragraph [0005] - paragraph [0020] * * paragraph [0026] - paragraph [0047] *	1-15	INV. G10G1/00 G10H1/00 G10H1/40 G09B15/02
A	ARZT A ET AL: "Automatic page turning for musicians via real-time machine listening", 18TH EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, GHALLAB METAL, PATRAS GREECE, 25 July 2008 (2008-07-25), XP001544133, ISBN: 978-1-58603-891-5 * abstract; figures 1-3 * * Sections 2, 3 and 4. *	1-15	
A	Arshia Cont: "Modeling musical anticipation: From the time of music to the music of time",  16 October 2008 (2008-10-16), XP055516200, ISBN: 978-0-549-90641-4 Retrieved from the Internet: URL:https://tel.archives-ouvertes.fr/tel-00417565/document [retrieved on 2018-10-17] * Sections 7.3 to 7.9 * * Sections 8.2 to 8.6 *	1-15	TECHNICAL FIELDS SEARCHED (IPC)  G10G G10H G09B
A	US 2014/358265 A1 (WANG JUN [CN] ET AL) 4 December 2014 (2014-12-04) * abstract; figures 1-5,9,12,13-18 * * paragraph [0040] - paragraph [0067] * * paragraph [0068] - paragraph [0080] * * paragraph [0094] - paragraph [0122] *	1-15	
The present search report has been drawn up for all claims			
Place of search <b>Munich</b>		Date of completion of the search <b>18 October 2018</b>	Examiner <b>Lecoince, Michael</b>
<p>CATEGORY OF CITED DOCUMENTS</p> <p>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</p> <p>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 &amp; : member of the same patent family, corresponding document</p>			

1  
EPO FORM 1503 03/82 (P04C01)

50

55

**ANNEX TO THE EUROPEAN SEARCH REPORT  
ON EUROPEAN PATENT APPLICATION NO.**

EP 18 38 2392

5 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.

18-10-2018

10	Patent document cited in search report	Publication date	Patent family member(s)	Publication date
15	US 2017110102 A1	20-04-2017	CA 2950816 A1 CN 107077836 A EP 3155608 A1 FR 3022048 A1 FR 3022051 A1 JP 2017519255 A US 2017110102 A1 WO 2015189157 A1	17-12-2015 18-08-2017 19-04-2017 11-12-2015 11-12-2015 13-07-2017 20-04-2017 17-12-2015
20	US 2014358265 A1	04-12-2014	CN 104217729 A US 2014358265 A1	17-12-2014 04-12-2014
25				
30				
35				
40				
45				
50				
55				

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.*

**Patent documents cited in the description**

- US 7098392 B2 [0003]
- EP 2919228 B1 [0004]
- US 8530735 B2 [0006]
- US 6156964 A [0007]
- US 20050015258 A1 [0007]
- US 9280960 B1 [0007]
- US 9747876 B1 [0007]
- US 20010023635 A1 [0007]
- US 8660678 B1 [0008]

**Non-patent literature cited in the description**

- **M.F. MCKINNEY et al.** Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms. *Journal of New Music Research*, 2007, vol. 36 (1), 1-16 [0006]
- **HAYTHAM FAYEK.** *Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between*, 21 April 2016, (<http://haytham-fayek.com/2016/04/21/speech-processing-for-machine-learning.html>) [0046]