Method for acoustic object-oriented analysis and note object-oriented processing of polyphonic sound recordings

EP 2099024 A1

Publication number	EP2099024 A1
Publication type	Application
Application number	EP20090001575
Publication date	Sep 9, 2009
Filing date	Feb 5, 2009
Priority date	Mar 7, 2008
Also published as	DE102008013172A1, DE102008013172B4, US8022286, US20090241758
Inventors	Peter Neubäcker
Applicant	Peter Neubäcker
Export Citation	BiBTeX, EndNote, RefMan

Patent Citations (5), Non-Patent Citations (5), Referenced by (4), Classifications (7), Legal Events (5)

External Links: Espacenet, EP Register

ABSTRACT

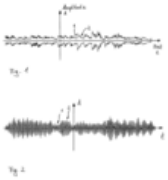
The method involves reading out the time signal section-wise by a window function and an overlapping window. The Fourier transformation of the readout signal is executed in a frequency chamber, particularly by executing a discrete Fourier transformation. An energy value for each bin is calculated from the frequency amplitude, which results from Fourier transformation, particularly by squaring the real and imaginary parts or creating a derived energy value. An independent claim is included for a computer program with a program code for executing the method.

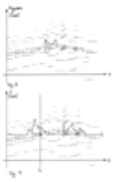
IMAGES (7)

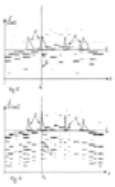

Search Report (1)

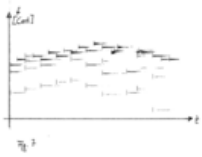

Search Report (2)


Search Report (3)









DESCRIPTION translated from German

[0001] The present invention relates to a method for sound object-oriented analysis and object-oriented processing of note polyphonic sound recordings according to claim 1,

[0002] It has long been known to undergo recordings with musical content of a sound post. Found this earlier in the studio engineering years even with the aid of expensive hardware instead, such as with complex filter banks, they are to currently used computers and special computer programs that are far less costly and therefore have found a larger distribution. The rest is to this development has done the advent of digital recording technology. The aim of such rework is usually to improve the sound of the recordings or incorporate sound effects. This kind of post tonal in common is that they work in and the effect-oriented musical content of the signal is not able to recognize, but the audio signal is understood only as a signal in the time-varying amplitude.

[0003] In the prior art such as are disclosed in EP 0750776 B1 or DE 696 14 938 T2 A method and a device for changing the pitch of the sound, and audio signals is known. It is considered disadvantageous that there can not be worked with complex program, but as it is at a normal music production.

[0004] It is desirable to manipulate audio material at the level of single notes, from which the sound recording. It is known in the art to extract from an audio recording individual notes with regard to its grade level, note length, and time of occurrence. One such example is from the note extraction DE 10 2004 049 477 A1 is known to determine a melody line of an audio signal. From WO 02/084 641 A1 is known in a note-based

CLAIMS (31) translated from German

1. Process for sound object-oriented analysis and object-oriented processing a note polyphonic, digitized sound recording, (t A,) as a time signal F is present, including analysis and the following processing steps:

   - Reading of the time signal portion-wise F (A, t) using a window function and overlapping windows,

   - Fourier transformation of the read-out signal into the frequency domain, in particular by performing a discrete Fourier transform,

   - Calculating an energy value E to each bin of the frequency amplitude, which results from the Fourier transform, in particular, by squaring the real and imaginary parts or forming a derived energy value

   - Generating a three-dimensional function F (t, f, E),

   - Identification of Event objects,

   - Identification of objects and conditions

   - Comparison of the temporal occurrence of event objects and note objects and assignment of event objects to objects and conditions in the event temporally plausible occurrence

   - Calculating spectral factors to share each note object

   - Associating of signal components of the frequency signal F (t, f, e)

description to reference the audio signal in a database, the conversion of an audio signal. A processing of the extracted notes, for example, by changing the frequency or moving in time, will not take place. These documents further prior art is mentioned.

[0005] A particularly critical point in editing of audio material is that the original sound impression as a voice to be retained even after processing. This is done in an excellent manner in which belonging to the state of the art software "Melodyne" of Celemony Software GmbH, which is based on a touch-based approach. However, this software requires that material be unanimous. Chord instruments like guitar, piano or chorus can not edit tonbasiert been satisfactory. Previously such recordings could only chord chord as cutting or time-stretching or pitch in time, but without access to individual notes in a chord. So it was not possible to use a single chord change (eg, the E a C major chord) (eg It is C-minor), without at the same time to work with the other notes of the chord.

[0006] Chord recognition and approaches to the identification of individual scores in the prior art, however, already known to be able to print such notes (WIDI software) or to identify songs automatically (DE 10 2004 049 477 A1).

[0007] It is the object of the present invention provide a way can be made as a note object-oriented processing of polyphonic material. This object is achieved with a method according to claim 1. Advantageous embodiments are specified in the dependent claims.

[0008] The method of the invention identified in a first step in an automatic musical material recorded audio objects in the sense of notes. These objects are then extracted sound in a second step from the total uptake and thereby rendered manipulated without noticeable loss of sound or sound noticeable distortion. This sonic extraction is not found in the prior art, meaning that an identified object is assigned a rating of the overall sound. This may even be the original recording music freely modified in such a way that the score can change their relative position to each other in pitch and time, but the original sound expression intact. The method further provides that the identified note objects are provided to the user for editing. The user can then change one or more musical objects, eg move in pitch. After such processing by a user, the tonal reproduction, re-synthesis is performed by the modified object is reproduced along with the unmodified objects or with the reduced overall signal to the changed object. The starting material may be present already in digital form, or an analog signal. Prior to the analysis of an analog signal digitization must then be carried out if necessary.

[0009] The method may find various applications. It is possible to eliminate such single wrong note specifically: a pianist accidentally struck a tone too much, which is removed in post processing. Another application concerns the retuning, so the correction for a detuned guitar or an unclean string section. Tempered recordings can be transferred to a pure atmosphere. It can be Umharmonisieren recordings, such as the guitar riff from a C major in F minor. Until now, shifted in pitch a chord only collectively but not the harmonious relationship of its individual sounds are changed. The possibilities range even up to a new composition, as there is access to the individual notes.

[0010] The inventive method is aimed primarily at individual tracks recorded during music production. But it is also applicable to ready-mixed tracks to make them appear in a new musical guise.

[0011] With the above, the prior art was feasible only for monophonic sound material isolated for processing, for example, recorded vocal or wind parts, where only one note each with an identifiable pitch will sound ie. With the inventive method, the aim of the individual note editing is feasible for polyphonic played instruments, ie those in which usually several notes or entire chords played simultaneously, such as piano, guitar, etc., and not just the chord as a whole is transposable (ie change the pitch while maintaining the relative pitch relations within a

on the detected object based on the calculated percentage score factors

- Reverse transformation of a music object associated frequency signal components in a time signal,

- Graphical representation of the objects and conditions and / or event objects in a time / frequency display on a monitor,

- User-driven or automated processing of one or more objects and conditions,

- Storage of the time signals processed note objects

- Reproduction of recorded time signals processed note objects along with the object to which a staff time associated with decreased signal time signal.

2. A method according to claim 1, **characterized in that** in the function F (t, f, E) is a function F '(t, f, E') is calculated for a with energy values E ', the t is the sum of all energy values E at a time a fundamental frequency f and all multiples thereof.

3. Method according to claim 2, **characterized in that** the energy values at the multiples of the fundamental frequency on the weighting added to a factor other than 1.

4. The method of claim 2 or 3, **characterized in that** for the identification of objects and conditions the following steps are performed:

   - Determining a maximum energy in the function F '(f, t, E'),

   - Determining a belonging to the maximum of values related field

   - Assignment of field values found for each object a grade.

5. A method according to claim 4, **characterized in that** the energy values E of the object values assigned to the score field of the function F (t, f, e) are subtracted.

6. Method according to claim 5, **characterized in that** the energy values E only in the height G * E are subtracted, wherein the gain G 0 <G <1.

7. Method according to claim 6, **characterized in that** the factor G is a user-variable parameters.

8. A method according to claim 5, 6 or 7, **characterized in that** the maximum search is continued at a function of the energy values are removed or. At a calculated therefrom function

9. Method according to one of the preceding claims, **characterized in that** the maximum search is carried out iteratively until a termination criterion.

10. The method of claim 9, **characterized in that** a total energy value E_{tot} for function F (t, f, E) is calculated, and the iteration is terminated as soon as a certain proportion H * E_{tot} this total value has been assigned to the detected note objects, especially when this proportion exceeds 90%.

11. The method of claim 10, **characterized in that** the factor H is a user-variable parameters.

12. Method according to one of the preceding claims, **characterized in that** an identified object in a subsequent automated Rating

chord), but also and especially the notes within a chord sound relative to each other can be changed, such as a change of major to minor chord sound.

[0012] The method relates to priority already recorded musical material, and not to analyzing and working in "real time", ie the moment of musical performance. This has to do with the fact that the inventive method, a meaningful recognition of "notes" implies in terms of completed objects in a certain context. Therefore, although the analysis can be carried out at a current recording in the background, but it has to be accessed from several seconds to a previously recorded time in the context of such magnitude.

[0013] The method is designed to search for a particular pitch and pitch waveforms in the total signal, and to distinguish, but not to isolate individual sound sources. It is not a question in a room separate from eg a recording of the sounds of a street or multiple speakers each source of noises or voices. There are usually two notes that were played on the same pitch at the same time by two different instruments, identified as a single object, as the sound of many first violins in the orchestra, playing the same note, as a single note. The term of the note object mainly used here emphasizes that "scores" for the purposes of this invention is not actual notes in a musical sense must be meant, although an unidentified object touch of a note can correspond in the actual musical sense, but need not be.

[0014] To-find notes must not, such as a piano, have a time-constant pitch, but can also, for example, over time have a voice with vibrato and / or portamento any change in pitch. If such pitch curves are the notes remain consistent, recognizable in time tracking of the signal. To this extent, two tones, which intersect at a different course in pitch when two different notes are identified.

[0015] The method essentially consists of two main steps: a) identification of participating in the overall sound of individual objects, so the notes and possibly also the more Fatos / percussive sound events, and b) sound resolution of the overall sound into the individual objects found whose sum gives the overall sound, However, each of which is characterized separately manipulated without affecting the sound of the other objects, or the overall sound in the sense of unwanted audible artifacts. Particularly in the step b), the present invention differs from the prior art.

[0016] The result of the identification step a) can not) be used even taken in the sound of separation step b, if it's all about knowledge and representation of musical content, without the sound to be encroached themselves. This may be the case if one wants to produce a musical notation of existing tracks or want to illustrate the musical content in other ways graphically. Or it can be used to identify the sequence of the musical harmonies and name it, for example to enrich the music with other instruments.

[0017] The quality of the result of identification step a) also affects the quality of the sound resolution in step b). If step a) only as mentioned is used to determine the musical content, it is usually sufficient to determine that at a certain time a sound object occurs, the example, the pitch of the note "C sharp" was, and how long sounds of this object. Conversely, if the sound resolution in step b) is the goal, then it is advantageous as much information as possible to evaluate the progress and the parameters of the note found objects, such as the exact course of the pitch curve in time, the amplitude of the object and its change in time, the nature of the onset of a note, the consistency of the mixture in proportion to its partial tones with similar pitch notes in the same recording, etc.. Depending on the claim but you can also dispense certain information.

[0018] As described, the main feature of the audio-to-find single material objects that they have a consistent pitch or pitch a consistent course.

cleanup is discarded if there is one or more of the following criteria:

- The energy of the detected object score is compared to the total energy is very low,

- The pitch and amplitude characteristic of the note object substantially coincides with that of another object and conditions lower frequency, particularly if the lower frequency is used as the fundamental frequency in question,

- The frequency of one distance to the other object Rating Rating objects is very large.

13. Method according to one of the preceding claims, **characterized in that,** in a post-processing step, a user disconnects automatically identified objects and conditions, connects and / or eradicated.

14. Method according to one of the preceding claims, **characterized in that** calculation to each bin an instantaneous frequency from the phase differences of adjacent bins is determined that as the frequency of the bin in the function F (t, f, E) or F '(t, f, e ') is used.

15. Method according to one of the preceding claims, **characterized in that** for finding the event objects to each bin a tonality value and / or Geräuschhaftigkeitswert is calculated.

16. A method according to claim 15, **characterized in that** the energy values of each bin are weighted with the tonality value.

17. Method according to one of the preceding claims, **characterized in that** use is made for the identification of objects and conditions on deposited notes.

18. Method according to one of the preceding claims, **characterized in that** the detected event objects from the time signal F (A, t) can be extracted, and the separation takes place on the remaining sound signals.

19. A method according to claim 18, **characterized in that** the event objects are stored separately.

20. Method according to one of the preceding claims, **characterized in that** the spectral component of a grading factors object from a stored function spectral component can be calculated.

21. A method according to claim 20, **characterized in that** the stored spectral component function is the mathematical model of the image of a sound note object that establishes one or more of the following postulates:

- There is a tone with a fundamental frequency corresponding spectral components on integer multiples of the fundamental frequency, namely overtones on harmonic frequencies,

- The amplitude curve of the harmonics to a fundamental follows a law which is derived from one or an empirically determined Obertonmodell Obertonamplitudenverlauf,

- The Obertonamplituden stand in a fixed ratio to the course of the fundamental energy.

22. A method according to claim 20, **characterized in that** the stored spectral component function is the mathematical

Furthermore, they should have a consistent course of their individual partials. That is, it is assumed that a to-find sound object from the course consists of a tonic, as well as for an unlimited number of harmonics to be the approximate integer multiples of the fundamental frequency. Further, the history object associated with a sound partials should have no cracks unmotivated. These assumptions are derived from the properties of the generated sounds of natural instruments. Therefore, the inventive method may have limited in cases where music was produced synthetically in any arbitrary manner.

[0019] To be identified in the audio material and objects described above can be seen as "sonorous" refer to objects in the claims they are named as a note objects. Their main property is to have a pitch or a pitch during a perceptible duration of their time and the waveform signal behaves essentially periodic or quasi-periodic. It will be of the non-sonorous objects differed, the noise objects. Event covering objects are a subset of the noise objects.

[0020] The event objects are characterized in that they have a sudden increase in amplitude in the time signal and for this reason are not periodically at this point. Also, you usually sound out quickly. These objects are usually musically produced by a sudden event, such as by striking or plucking a string, or the beating of a percussion instrument like a drum. In this context, they may also be further distinguished from the process of the invention: if follows an event-like object immediately increase the amplitude of a sonorous object, it can be assumed that the event represents a note's velocity, and therefore can be assigned to this note. In this case, the sonorous and the event object can be grouped for display and manipulation of a single object, which is also classified as a central object. For producing sound in the resynthesis, ie the sound reproduction after editing a note object or multiple objects touch, but it may be useful to separately process the two mentioned subobjects sound reasons.

[0021] If the event object can not be assigned to a subsequent sonorous object, it can be assumed that it is a purely percussive note with no pitch rhythmic understood meaning, such as a drum beat. This can be handled in the further processing therefore different.

[0022] The noise objects have in common with the eventful objects that they do not have a periodic sound content, but differ from them in that they do not start suddenly and does not fade quickly, but can be temporally extended. Such objects may be, for example sibilance of consonants in singing, breath sounds, the player background noise such as a finger slides on the guitar string, or not at all to the actual music part of noise. Such as noise objects could be provided to the user separately available for further manipulation or simply flat will be treated as "residual signal".

representation of the sound of a play on a particular instrument and conditions.

23. The method of claim 20, **characterized in that a** plurality of spectral component functions are stored, in particular for several instruments different unit functions, in particular those according to claim the 22

24. A method according to claim 23, **characterized in that** one of said plurality of content by the user functions is selected.

25. A method according to claim 23, **characterized in that** a music object one of the plurality of content features is automatically assigned by the user is detected when input or automatically, on which the note was played instrument ..

26. Method according to one of claims 20 to 25, **characterized in that** the spectral component function is a window function performs a weighting in the frequency range having a predetermined frequency width.

27. . Method according to one of the preceding claims, **characterized in that** a residual signal is calculated by subtracting all the objects and event objects, note assigned time signals from the original time signal

28. A method according to claim 27, **characterized in that** the residual signal to a further identification of notes or event objects is subjected.

29. Method according to one of the preceding claims, **characterized in that** for the sound reproduction according to the sound processing of a portion of the note object note object is subtracted from the total sound, and the resulting difference signal is played back together with the sound of the note content object.

30. The method of claim 27, **characterized in that,** in the sound reproduction according to the processing of an object by a user score the residual signal is to be played.

31. Computer program having a program code for performing the method according to any one of the preceding claims when the computer program runs on a computer.

[0023] The method of the present invention is illustrated by way of example with reference to the figures, in which results of individual process steps are shown. The procedure was performed on a 6-second excerpt from a piano recording of the Invention no. 1, C Major by JS Bach. In the drawings:

Figure 1
: the audio signal F (A, T) as the amplitude A over time t;

2
: an enlarged detail of the time Figure 1 about 0.5 seconds;

3
: after transformation of the audio signal of the Figure 1 in the frequency range grautonkodierten the energies of the individual bins in their Augenblicksfre sequence, with dark areas indicate high energy, in a representation of the frequency f in cents over the time t: F (t, f, e);

Figure 4
: a view according to 3 with a cut in the x and y-direction, which shows the calculated energy E at this frequency f_1 and the time t_1;

Figure 5
: an energy landscape F (f, t, e) similar to that in Figure 4 shown, in which the summed energy values of a

window function, and have been smoothed in the time direction, a section in x-and y-direction at the frequency f_1 and the time t_1;

Figure 6

a relevance to the energy landscape of the countryside Figure 5 with a cut in the x and y direction with the first detected peak, and

7

Mark all found objects after a iteratively executed identification.

[0024] There are terms used to explain the invention to be defined first in importance. The terms "time domain" and "frequency" corresponds to the usual use of time domain and frequency domain in English, and therefore relate to investigations or calculations either in the original time signal F (A, t) itself (= the time domain) or in its representation F (f, t) in a discrete Fourier transform, FFT specifically, transferred form (= frequency range).

[0025] "Window functions" are used to show and hide a signal at its processing on a time-or frequency location. You can refer to the time or the frequency spectrum, ie. The window shape is not fixed and can be a triangular window in a simple case. Better results are obtained as a by-Hann window. The choice of window shape can be optimized for the particular purpose. Overlapping windows are used for the hereinafter-described embodiment of a method according to the invention.

[0026] As "bins" are referred to the frequency bands, resulting from the FFT. This means that the frequencies may be changed a tape are referred to as the instantaneous frequency of the method is applied.

[0027] The instantaneous frequency of a bin is derived from the consideration of the phase value of each Bins. If the analysis windows overlap, can be made of the difference between the actual phase value of the present bin, the instantaneous frequency of the respective bin is determined by the expected progression of time and. The more the overlap, the more adjacent bin may represent a particular frequency does not have to coincide with the theoretical frequency of the even bins.

[0028] It can be calculated for the inventive method for the energy-like quantities bins that receive the shorthand notation E and referred to as "energy" and "magnitude" to be called. After each bin of the FFT are respectively associated with an energy that, according to (Re Re *) + (Im * Im) calculated from the real and imaginary part of the Fourier series, so that the amplitude of the frequency related value occurs. The magnitude is the square root of it. For optimal scaling of the value of the magnitude value in the evaluation, but can be optionally raised to an appropriate value between 1 and 2, so that the size obtained is then something between power and magnitude. The concept of energy is thus used here in a general sense, and not in the sense of sound energy or other energy terms.

[0029] The cent is used in a musical context, a measure of frequency ratios, ie interval sizes, defined as cents = log (f1/f2) / log (2) * 1200th Therefore, in this unit, a musical interval is independent of its absolute pitch always the same, namely semitone = 100 cents, octave = 1200 cents.

[0030] For the identification of note objects to the method described uses metaphorically speaking of an energy landscape, among which a three-dimensional mathematical structures F (t, f, e) is understood in which the x-axis is the time t, the y-axis the frequency f and the z -axis represents an energy E at the respective time / frequency location. The frequency axis is scaled as appropriate cent to musical intervals in each frequency region are always equal. The energy landscape representing the method to be described further by means of discrete points, the time axis through the measuring points in time, for example, approximately 0.01 s interval, the frequency-axis by a distance of points in the interval, for example 5 per cent. In some steps of the discrete points are converted into continuous functions on window spectrograms, but this is to be regarded as optional.

[0031] The studied material is by recording audio and analog-to-digital conversion as an audio file in PCM format (Pulse Code Modulation, values and discrete-time signal). Size information in the text, such as for analysis window, refer to a digital signal that / s is present at a sample rate of 44100 samples. For other sample rates, the sizes were adjusted accordingly.

a) Identification of note objects and event objects

[0032] The method described exemplary works during the analysis as well as the sound extraction for specific sub-tasks directly in the time domain (1 and 2), For others in the frequency range (3 to 7). Here, the treatment, the treatment in the frequency range in the time domain is better for the event objects, better for the sonorous objects. Note there is searched for objects in the discrete time signal processing is not continuous, but it is in both the time domain and in the frequency domain is always between a stored time interval, which is then analyzed.

[0033] Figure 1 shows a waveform F (A t) to an audio recording. Figure 2 shows an enlarged detail of it. 1 with an event object is referred to in these figures that can be fixed to an amplitude jump. The area labeled 2 shows a clear contrast, more periodic course.

[0034] For the processing in the frequency domain, the signal is read out and uniformly successive overlapping window functions and a first converted to a complex FFT array for the respective time slice. The size of the FFT can be eg 2048 samples, the overlap should be at least 4-fold. This results in such distances as the time slices of 512 samples or about 0:01 sec

[0035] A further example of a transformation result is processed Figure 3 . It is there a function F (f, T, E) shows that the benefits derived from the Fourier transform of the in Figures 1 and 2 shown timing signal F (A, t) is characterized in

that the Bins and the instantaneous energy spectrum to be determined. As shades of gray encoding, the energies of the individual bins are applied in their instantaneous frequency over time. 3 Event objects are distinguished in this view by the uniform distribution of instantaneous frequencies, which indicates the noisiness of this object. Note objects are characterized by the concentration of power in a few instantaneous frequencies. This difference is due to the fact that a non-periodic waveform can be adequately represented by only a superposition of a large number of periodic functions can be periodic as well areas by a few periodic functions mapped. Therefore periodic signal areas in time lead to a strong overlap of adjacent energy bins 4 on a common instantaneous frequency.

[0036] To the signal F (f, t, e) in the frequency range, the following values were determined: the magnitudes of all bins, the instantaneous frequencies of all bins, the tonality values of all bins. The tonality value is a calculated value, which represents the degree of periodicity in the Binfrequenz. It is calculated for each bin by determining how close the instantaneous frequencies of the neighboring bins are located on the instantaneous frequency of the bins in question. The number of consolidated neighboring bins is equal to the number of windows overlap, since this determines how many bins may represent a frequency. The tonality value of a bin is the higher, the closer, the instantaneous frequencies of the bins in its environment together. Here, a high tonality value that tends to present a note object, while a low tonality value tends to speak for an event object. The tonality values are scaled to a range of values from 0 to 1. In addition, each bin is assigned a noisiness, which is derived directly from the tonality value, and is calculated as 1 - tonality value. A sudden increase in the noisiness implies an event object.

[0037] It is a landscape of energy is generated in the example Figure 4 as a section along the time t_1 and the frequency f_1 is shown, and t is the distribution of the energy E in the time and frequency f and represents the actual finding of note objects serves as ridges in the landscape. Smoothing in the time direction may be carried out. Furthermore, a preference for the further calculations by displaying the energy values over a window function of the energy landscape of Figure 4 recovered modified energy landscape used. This then results in a Figure 5 illustrated smoothed function F (f, T, E). In both characters step 5 objects clearly, get those high energy values.

[0038] Since it first comes to finding sounded-like objects, the calculated energies of the bins are additionally weighted by the determined tonality values: for each time slice, its energy is multiplied by its tonality value for each bin. This made weighting changes the result only in degree, can therefore be omitted. According to the instantaneous frequency of the bin then his penny-position (= y-position) is determined in the landscape and starting summed up by this point, the energy tonality product with a certain distribution width in cents towards a window function to the landscape. The width of the distribution window in cents is useful as on the order of a semitone. The representation of Figure 4 was such a weighting with the underlying tonality values already. If all time slots are summed up in the landscape, the landscape can be smoothed in the time direction with a low-pass filter (see Figure 5). This finding related note objects is easier than ridges. The total energy of the landscape is summed and represents a value for the termination criterion of the subsequent iteration available.

[0039] Identifying the note objects is done through an iteration in the manner that you have the most pronounced salient object is tracked in terms of the maximum of the ridge and its energy is then subtracted from the landscape, and then the next most obvious salient object is searched, etc. The but sonically most clearly protruding object is not identical to the highest mountain range in the energy landscape F (f, t, E). This is due to the fact that one sounded bad object is not defined by a single ridge in the energy landscape, but it must be assumed that the energy in the mountain ranges are the integer multiples of a fundamental frequency, even unto this very sought sonorous object because they overtones the root to the base frequency. It may well be that the energy of the fundamental is weaker than that of the higher partials, but still want the object to be found on the frequency of the fundamental and tracked. Of these considerations is that, although the fundamental frequency determines the pitch of a sound, the sound of a tone but is strongly influenced by the harmonics.

[0040] To take this into account is metaphorically a second landscape created called relevance landscape, the example in Figure 6 is shown, and their x-and y-axes, and the size of the energy landscape of F (f, T, E) is the same as deriving its z-values E 'but from this: to, for each x, y coordinate point of relevance landscape F '(f, t, E') as z-value E 'is the sum of all values of z e formed in the energy landscape F (f, t, e) are in this xy point and at all points, the the integer multiples of the output frequency point correspond. It is useful to add, with increasing atomic number of the multiple frequencies available energy value with decreasing weight. In this way, a landscape Relevance F '(f, t, E'), in which the possible root notes are considered with their partials, and the highest point of the sound most relevant point of note is most relevant. In the Figure 6 relevance landscape shown F '(f, t, E') shows the same energy protruding objects 5 It has been through the consideration of the energy contained in the harmonics to the Figure 5 a shift in the relative energy levels result.

[0041] The discovery of the central objects in the landscape Relevance generated as just described, which is basically a special energy landscape, namely the Obertonenergie considered done by an iterative mathematical process. It is explained below how the objects are finding the note in this relevance landscape, without the invention is limited. After finding the note objects could in principle be carried in one or more of the other above-mentioned modified energy landscapes, but would have the disadvantage that the overtones would be identified as separate notes and would be connected by post with the fundamental tones for example. The task of separating note objects of sound, it is then solved well when the link between fundamental and harmonic tones succeed. Therefore, the maximum search in the relevance landscape is preferred because it leads to the best results.

[0042] It is the highest point of the relevance landscape is first sought. In Figure 6 was found at t_1 and f_1 the energy maximum. The ridge of belonging to this maximum ridge is followed forward and backward in time direction. To the adjacent time slice of the landscape that is sought the last point found in pitch direction closest peak in the landscape. If the distance to the nearest maximum is so great that a continuation of the pitch line is not plausible as the same object, eg a jump of more than 50 cents of a time slice to the next, the search in the current direction canceled. Similarly, the search is terminated when a certain amount of the maximum value found is not met, such as 10% of the initial value. The persecution of the ridge is preferably done in the relevance landscape, because the course is more in Mitgewichtung result of the change in pitch of the partials of the object sought. When the search is stopped in both directions, a new note object is created and added that all the points of the ridge as its found in pitch accordingly. In the claims in this connection it is stated that a maximum belonging to the value field is determined. The method of determining this value field can be other than described above take place, for example by other mathematical methods are applied. It could, for example in all directions away from the maximum point by point the value field are queried until a threshold value in each direction is maintained. All points lying above the threshold would be assigned to the maximum as the value field.

[0043] Then the energy landscape E (f, t, e) deprived the calculated energy of the detected note object, namely at the sites of the course as well as its fundamental note of all the partials, ie the integer multiples of the fundamental frequency. In principle, this also in the relevance landscape E '(f, t, e') place, but this would be a worse variant, because the possible superposition of harmonics belonging to different basic tones, with removal of energy from the energy landscape will be better off.

[0044] Advantageously, but not the entire available energy is extracted from there, but only a fixed percentage, say 50%. This proportion can be set as a parameter by the user to other values, because depending on the audio material other units may give better results. With strong overlap in the overtones as may lead to a reduction of 25% to better results. The withdrawal only a portion of the energy is useful because first is not known if others do not have the same sounding note objects partials that are close to the partials of the note object found first. By the only partial power hood another note objects can be found in the following iterations yet.

[0045] Reducing the power to the places of the given frequency energy landscape E (f, T, E) is advantageously in the form of a back in the frequency direction ausblendenden up and down window function whose width is of the order of a semitone. If a model of the overtone spectrum of the sound is known, for example, because to the sound-producing instrument there is a reference spectrum is modeled or known energy deprivation in the overtones of this reference spectrum or model can be accordingly. The note object "remembers" the amount of energy withdrawn by them for their subsequent evaluation by their proportion of this energy is attributed.

[0046] In the time domain, which is affected by the newly found note object, the relevance landscape is recalculated as described above, since in this time range, the energy landscape has changed as the basis for relevance landscape by extracting the energy.

[0047] For the newly found note object determines whether it overlaps in time and in the course of its fundamental frequency with another object previously been detected note. If so overlaps with such or because it directly connects, it is plausible that it is the same note object, it is added to this (possibly with extension whose pitch curve). Otherwise, it is accepted as a new note object in the amount of the note found objects. As for example only 50% of the energy for the note object are extracted in each iteration, as a rule, each note object is found several times in the course of the iteration.

[0048] The iteration is continued by again the highest point is sought relevance in the changed landscape. The iteration is continued until a termination criterion. A favorable termination criterion for the iteration is the reduction of energy in proportion to the original energy in the energy landscape. The iteration can be stopped eg when only 10% of the original energy in the energy landscape is present. This can also be provided to the user as to changing parameters.

[0049] The discovery of the event objects, which are characterized by a sudden increase in the noise portion of the signal, can be carried out either in the time domain signal by the rise especially high-pass filtered signal components is monitored, or in the frequency domain using the noisiness of the bins to weighted the energies of the respective bins are summed for each time slice. In both cases we obtain a trajectory of noise portion of the total signal. At the points of greatest slopes of these curves, possibly defined by a slope threshold, the event objects are assumed.

[0050] Found in the previous step event objects can either occur in isolated form for himself in the signal, as will be the case in purely percussive events, or they can be the attack sounds the note objects that have been previously found in the iteration, as in plucked or battered tonal instruments such as guitar, piano, etc. will be the case. To distinguish this, is examined for every detected event object at the time of its occurrence, whether immediately after the occurrence of the event occurs, a significant increase in their energy in one or more of the touch objects located there. If this is the case, the event object is understood as a stop for the note object and assigned. If the energy increase is the case with several notes, the event object is associated with all these notes. If the energy increase occurring in a note object in the middle, the note object where it is separated and from there understood as a new note object. At the time of the eventful object if no corresponding note object is found, the event object is understood as an isolated percussive event. Figure 7 shows the central objects found in the present example, together with the vertical lines marked as eventful objects that could be assigned to this note objects.

[0051] Finding the note-like objects should follow with a benefit assessment step. When searching for note objects in the described iteration usually more objects are found, as are musically plausible available. Therefore, the amount of the note found objects is checked at the end even after various plausibility criteria and, if not removed sufficiently plausible note objects. A plausibility criterion as the relative energy and the occlusion. Generally be found in the described iterative procedure to score many small objects with too little energy. Therefore, it is investigated how much energy has the note in relation to the total energy in their time domain. If it has not enough energy relative, it can be removed.

[0052] Sometimes objects are identified as separate notes, which are actually overtones of another existing note. Here, for example, can be examined if the higher note has a separate course in pitch, amplitude and duration, or whether it essentially runs as a lower note in these parameters. If the latter is the case, the object can be removed or the lower note is struck.

[0053] More reviews can be made a musical perspective. For example, if a note object is in their pitch very isolated area (very high or very low, if there are no other notes), it is unlikely musical. For example, if a note is combined with other notes in its correctly tuned and temporal proximity to an ascending or descending line, it is musically very likely, even if she is otherwise rather weak, etc. All these criteria can be depicted mathematically as weight, to arrive at a plausible possible amount of note objects.

[0054] This identifying step can be described also follow intervention by the user, the retrieved note objects are graphically represented in a suitable manner, for example in the in 7 type shown, for example, and the mouse and / or menu driven as a note identified objects can not divide or separate notes can unite to an object. The user can also delete individual items or add more items to consider. To the objects it can be offered for activation that were previously evaluated in the automatic evaluation than less relevant.

[0055] The automatic identification can optionally be optimized by the notes of the recorded piece of music is stored, so that in the above method tries specifically because of the stored notes, there find basic sounds that correspond to the frequencies of the stored notes. This can be done by evaluating a stored MIDI file containing the notes of the composition taken as. It may alternatively be taken simultaneously with the recording of the total signal support tracks actually used, eg. Einzelmikrofonierung by the instrumentalist or singer involved, or a guitar by individual acceptance of the strings In the monophonic signal then pronounced the desired notes of the individual votes of the total signal can be clearly identified and separated thus better overall sound for editing.

b) assigned to note sound objects

[0056] After the identification of the individual objects involved in the overall sound is complete, the sound resolution of the overall sound can be done in a next step. The accurate determination of the existing note objects, their pitch and amplitude curve and the nature of their onset is a factor influencing the quality of the result of sound resolution. When sound resolution described below, the overall sound is only so far broken down into individual sound objects, as it is necessary for resynthesis of the new overall sound. If so will only touch a single note found in a complex chord and changed by the user in their pitch, for example, then only the signal that has extracted a note and subtracted from the original signal. There are thus produced the more individual signals, the more notes will be changed. Each is then respectively a monophonic-periodic signal and can be changed, added up and played with already known methods for time-and pitch-independent playback.

[0057] In a first step the separation of the sound event objects from the original signal are extracted. If the original signal is decomposed into belonging to the individual note objects individual signals, while the eventful places are smeared in the time signal by dividing the frequency spectrum. Therefore, it is useful first to separate the locations of the event objects from the time signal and perform the decomposition into the individual signals of the central objects of the residual signal thus produced. However, this optional step may also be omitted.

[0058] For the separation of event objects, a modified time signal is first generated, in which the sonorous parts are removed as completely as possible. For this, the magnitudes of all bins are multiplied by the Geräuschhaftigkeitswerten the bins and recreated with the new mag and the original phases of the FFT time signal in the frequency domain. Optionally, you can enter another damping factors for the lower signal components in the factors for the magnitude, because often the higher interests are relevant to the event objects. For this new noise-like time signal is applied to the points at which event-like objects have been found in the identifying step as described above, the timing signal of the event objects separated out with an appropriate window function, such as a short rise time of approximately 0005 s, and a decay time of about 12:05 a.m. These short-time signals, see the event objects are removed from the original time signal and as saved separately.

[0059] There then follows a separation of the central objects of the original signal. The distribution of the original signal (of the eventful shares have been deducted) is happening in the individual sounds of the central objects in the frequency domain. To the original signal after its modification through the separation of event objects (see above) first re-transformed into the frequency domain.

[0060] The division into the individual sounds of note objects based on the fact that each note object in each time slice, a "claim" to a spectral component of the total signal F (t, f, e) logs. This claim is mathematically represented by spectral component factors are calculated for each note object from a spectral component function, for example extracted from a model of the sound of a single note. This model can easily be predefined, or it can be modeled on the actual sound of an instrument when the instrument is known to back the music object. The model in this

example is based on the following components: It is assumed that the fundamental course of the note object exist overtones as spectral components at integer multiples of the fundamental frequency. It is further assumed that the amplitudes of the harmonics follow a fundamental one harmonics model. This can be the decrease in the amplitude of the reciprocal of the Obertonnummer in the simplest case. The harmonics model can, however, also reflect the Obertonamplitudenverlauf which has been derived from an empirical sound. It is finally assumed that the harmonic amplitudes are in a relationship with the course of the fundamental energy of the note object. This in the simplest case it can be assumed to be proportional to the tone power, but it can also be another relationship can be derived from an empirical sound.

[0061] Based on these assumptions, a spectral component function is determined, can be different for example for different instruments, and it will calculate the spectral component factors for each note object in each time slice, so the claims that this object has in each bin.

[0062] If more than one share functions are stored, for example can be provided that the user selects a portion of the functions. It can also be an automatic choice, such as when the user enters which instrument the note object has been played, or if it is automatically recognized that the grading object was played on a particular instrument, which for example may result from the fact that the Obertonamplitudenverhältnisse a note object Share correspond to a stored function.

[0063] The amount of the calculated proportion depends on factors Teiltonfrequenzen and the amplitudes of the partials, for example, following from the underlying model of the sound of a single note. Further, the amount of the proportional factors depends on the proximity or distance Teiltonfrequenz to the instantaneous frequency of the respective bin. The amount of the proportional factors in dependency of the distance can, for example enter via a weighting curve in the frequency range which should be wide enough such that even slight variations can be allowed by the given frequency. On the other hand, the weighting curve in the central region should be so narrow that the Obertonanteilsfaktoren of different simultaneously sounding notes of varying pitch of the overtones good enough the correct note can be separated and assigned. An appropriate weighting curve to evaluate the removal of frequencies can be eg a since elevated to the fourth power of Hann window, which corresponds to its total width eg two semitones.

[0064] If in the current time slice, all registered there as a sounding note objects found their claims on the proportional factors in all bins, the sum of the proportional factors all scores for each bin is normalized to 1. For each note object its own signal with the time duration of this note object is created. For each time slice, the magnitudes, or other suitable energy value of all the bins in accordance with the exemplary normalized proportional factors to note objects are divided. These shares of each note objects in the frequency domain are transformed back to the original phases of the FFT in the time domain and accumulates the time signals on the individual time signals of the central objects.

[0065] Since the Magnitudenanteile or other energy-like components have been previously altered, the ends of the signal in the time windows are no longer disappear after re-transformation into the time domain to 0, which leads to undesirable artifacts. Therefore, the result of the inverse transformation should be windowed again. Therefore it is desirable to take the square root from the values of the actually provided window function and then apply the window to the front of the FFT and inverse FFT.

[0066] Finally, the central objects with the eventful shares will be merged. Next up was already described that the objects touch the event objects have been assigned, and that a time signal was generated for the event objects. Now you can touch the objects to which event objects have been assigned, this time signal are added together at their beginning. If an eventful several central object objects have been assigned, because it was assumed that this was posted at the same time, the time signal of the eventful object is distributed in amplitude to the assigned rating items. This can be done in the appropriate ratio of the energies of the central objects themselves or on account of the assumed model Instruments ..

[0067] Event covering objects, which do not touch objects were assigned to stand with their extracted time signal as an independent percussive objects.

[0068] If you have time signal was generated for all objects detected note and their associated event objects, this time signals all scores are subtracted from the original signal. Because underlying sound model was provided in the sound distribution that the central objects consist essentially of partials, the approximate integer multiples of a fundamental frequency, not the entire sound and thus not the total time signal will have been distributed to the individual objects. Therefore, by deducting the timing signals of all the individual objects of the remaining original signal a residual signal which contains the more noise-like components. This residual signal in the resynthesis can easily be played with, or it may be the user, as a whole or cut into a further single objects that are made available for further processing.

[0069] The time sequence of the individual process steps mentioned above can also be selected differently. For example, the assignment of event objects to touch objects can also be performed immediately prior to resynthesis. This applies analogously to other process steps, such as identifying the event and touch objects or calculating the proportion of factors.

PATENT CITATIONS

Cited Patent	Filing date	Publication date	Applicant	Title
EP0750776B1	Jan 18, 1996	Sep 5, 2001	Ivl Technologies Ltd.	Method and apparatus for changing the timbre and/or pitch of audio signals
DE69614938T2	Jan 18, 1996	Apr 25, 2002	Ivl Technologies Ltd	Verfahren und vorrichtung zur änderung des klanges und/oder der tonhöhe von audiosignalen
DE102004049477A1	Oct 11, 2004	Apr 20, 2006	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.	Verfahren und Vorrichtung zur harmonischen Aufbereitung einer Melodielinie
US6057502 *	Mar 30, 1999	May 2, 2000	Yamaha Corporation	Apparatus and method for recognizing musical chords
WO2002084641A1	Apr 4, 2002	Oct 24, 2002	Brandenburg Karlheinz	Method for converting a music signal into a note-based description and for referencing a music signal in a data bank

* Cited by examiner

NON-PATENT CITATIONS

Reference		
1	*	CHRIS DUXBURY, NICOLAS CHÉTRY, MARK SANDLER, D MIKE DAVIES: "Efficient Two stage implementation of the Harmonic Maching Pursuit" EUSIPCO 2004, [Online] 2004, XP002533839 Gefunden im Internet: URL:http://www.eurasip.org/Proceedings/Eusipco/Eusipco2004/defevent/papers/cr1814.pdfs1VF1g> [gefunden am 2009-06-24]
2	*	KLAPURI A P: "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, Bd. 11, Nr. 6, 1. November 2003 (2003-11-01), Seiten 804-816, XP011104552 ISSN: 1063-6676
3	*	M. EVERY AND AL.: "Separation of synchronous pitched notes by spectral filtering of harmonics" IEEE Transactions on Audio, Speech and Language Processing IEEE USA, Bd. 14, Nr. 5, September 2006 (2006-09), Seiten 1845-1856, XP002533838 ISSN: 1558-7916
4	*	RÉMI GRIBONVAL ET AL: "Harmonic Decomposition of Audio Signals With Matching Pursuit" IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, Bd. 51, Nr. 1, 1. Januar 2003 (2003-01-01), XP011080322 ISSN: 1053-587X
5	*	VIRTANEN T ET AL: "Separation of harmonic sounds using linear models for the overtone series" 2002 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). ORLANDO, FL, MAY 13 - 17, 2002; [IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP)], NEW YORK, NY : IEEE, US, Bd. 2, 13. Mai 2002 (2002-05-13), Seiten II-1757, XP010804234 ISBN: 978-0-7803-7402-7

* Cited by examiner

REFERENCED BY

Citing Patent	Filing date	Publication date	Applicant	Title
EP2362375A1	Sep 3, 2010	Aug 31, 2011	Fraunhofer-Gesellschaft zur Förderung der Angewandten Forschung e.V.	Apparatus and method for modifying an audio signal using harmonic locking
EP2362376A2	Sep 3, 2010	Aug 31, 2011	Fraunhofer-Gesellschaft zur Förderung der Angewandten Forschung e.V.	Apparatus and method for modifying an audio signal using envelope shaping
WO2011104354A1	Feb 25, 2011	Sep 1, 2011	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.	Apparatus and method for modifying an audio signal using harmonic locking
WO2011104356A2	Feb 25, 2011	Sep 1, 2011	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.	Apparatus and method for modifying an audio signal using envelope shaping

CLASSIFICATIONS

International Classification	G10H1/06
Cooperative Classification	G10H2210/066, G10H1/06, G10H2220/116, G10H1/0008
European Classification	G10H1/00M, G10H1/06

LEGAL EVENTS

Date	Code	Event	Description
May 12, 2010	AKX	Payment of designation fees	**Designated state(s):** 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 SE SI SK TR
Apr 28, 2010	17Q	First examination report	**Effective date:** 20100326
Apr 7, 2010	17P	Request for examination filed	**Effective date:** 20100225
Sep 9, 2009	AK	Designated contracting states:	**Kind code of ref document:** A1 **Designated state(s):** 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 SE SI SK TR

Sep 9, 2009	AX	Extension or validation of the european patent to	**Countries concerned:** ALBARS