



(11) **EP 2 043 089 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
14.11.2012 Bulletin 2012/46

(51) Int Cl.:
G10H 1/42 (2006.01) G10H 1/00 (2006.01)

(21) Application number: **07117541.8**

(22) Date of filing: **28.09.2007**

(54) **Method and device for humanizing music sequences**

Verfahren und Vorrichtung zur Humanisierung von Musiksequenzen

Procédé et dispositif pour humaniser des séquences musicales

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IS IT LI LT LU LV MC MT NL PL PT RO SE
SI SK TR**

(43) Date of publication of application:
01.04.2009 Bulletin 2009/14

(73) Proprietor: **Max-Planck-Gesellschaft zur
Förderung
der Wissenschaften e.V.
80539 München (DE)**

(72) Inventors:
• **Hennig, Holger
37073 Göttingen (DE)**

- **Fleischmann, Ragnar
37081 Göttingen (DE)**
- **Theis, Fabian
93073 Neutraubling (DE)**
- **Geisel, Theo
37075 Göttingen (DE)**

(74) Representative: **Bach, Alexander et al
Mitscherlich & Partner
Patent- und Rechtsanwälte
Sonnenstraße 33
80331 München (DE)**

(56) References cited:
**US-A- 3 974 729 US-A- 6 066 793
US-B1- 6 506 969**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 2 043 089 B1

Description

[0001] The present invention relates to a method and a device for humanizing music sequences. In particular, it relates to humanizing drum sequences.

TECHNICAL BACKGROUND AND PRIOR ART

[0002] Large parts of existing music are characterized by a sequence of stressed and unstressed beats (often called "strong" and "weak"). Beats divide the time axis of a piece of music or a musical sequence by impulses or pulses. The beat is intimately tied to the meter (metre) of the music as it designates that level of the meter (metre) that is particularly important, e.g. for the perceived tempo of the music.

[0003] A well-known instrument for determining the beat of a musical sequence is a metronome. A metronome is any device that produces a regulated audible and/or visual pulse, usually used to establish a steady beat, or tempo, measured in beats-per-minute (BPM) for the performance of musical compositions. Ideally, the pulses are equidistant.

[0004] However, humans performing music will never exactly match the beat given by a metronome. Instead, music performed by humans will always exhibit a certain amount of fluctuations compared with the steady beat of a metronome. Machine-generated music on the other hand, such as an artificial drum sequence, has no difficulty in always keeping the exact beat, as synthesizers and computers are equipped with ultra precise clocking mechanisms.

[0005] But machine-generated music, an artificial drum sequence in particular, is often recognizable just for this perfection and frequently devalued by audiences due to a perceived lack of human touch. The same holds true for music performed by humans which is recorded and then undergoes some kind of analogue or digital editing. Post-processing is a standard procedure in contemporary music production, e.g. for the purpose of enhancing human performed music having shortcomings due to a lack of performing skills or inadequate instruments, etc. Here also, even music originally performed by humans may acquire an undesired artificial touch.

[0006] Methods exist for creating deterministic irregularities (offsets) between beats and the timings of notes: e.g. for playing back jazz music (swing) for example US3974729, US6066793.

[0007] Document US6506969 discloses how a musical sound sequence may be created by choosing rhythm randomly and applying interpretation effects for varying the volume and timbre of sounds.

[0008] Document GB2418054 details applying random variations to parameters of musical notes (amplitude, frequency of the note, length of time between notes etc.).

[0009] Therefore, there exists a desire to generate or modify music on a machine that sounds more natural.

SUMMARY OF THE INVENTION

[0010] It is therefore an object of the present invention to provide a method and a device for generating or modifying music sequences having a more human touch.

[0011] This object is achieved according to the invention of by a method and a device according to the independent claims. Advantageous embodiments are defined in the dependent claims.

[0012] The term sound to which the claims refer is defined herein as a subsequence of a music sequence. In some embodiments, a sound may correspond to a note or a beat played by an instrument. Each sound has a temporal occurrence t within the music sequence.

[0013] Preliminary results of empirical experiments carried out by the inventors strongly indicate that a rhythm comprising a natural random fluctuation as generated according to the invention sounds much better or more natural to people than the same rhythm comprising a fluctuation due to Gaussian or uniformly distributed white noise with the same standard deviation, even when using Gaussian instead of uniform white noise.

BRIEF DESCRIPTION OF THE FIGURES

[0014] These and further aspect and advantages of the present invention will become more apparent when studying the following detailed description of the invention, in connection with the attached drawing in which

- Fig. 1 shows a plot of a natural drum signal or beat compared with a metronome signal;
- Fig. 2 shows the spectrum of pink noise graphed double logarithmically;
- Fig. 3 shows a flowchart of a method according to an embodiment of the invention;
- Fig. 4 shows a block diagram of a device for humanizing music sequences according to an embodiment of the invention; and
- Fig. 5 shows another block diagram of a device for humanizing music sequences according to another embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0015] **Figure 1** shows a plot of a natural drum signal or beat compared with a metronome signal. Compared to a real audio signal, the plot is stylized for the purpose of describing the present invention, which only pertains to the temporal occurrence patterns of sounds. The skilled person will immediately recognize that in reality, each beat or note played is composed of an onset, an attack and a decay phase from which the present description abstracts.

[0016] The beats of the metronome occur on times t_1 , t_2 and t_3 and constitute a regular sequence of the form

$$t_n = t_0 + nT, \quad (1)$$

wherein t_n is the temporal occurrence or time of the n-th beat, t_0 is the time of the initial beat and T denotes the time between metronome clicks.

[0017] The human drummer's beats occur on times t'_1 , t'_2 and t'_3 and constitute an irregular sequence. The offsets o_i between the beats may be calculated as

$$o_n = t_n - t'_n. \quad (2)$$

[0018] Alternatively, the above definitions may also be generalized in order to track deviations of a sequence from a given metric pattern instead from a metronome. In other words, instead of taking regular distances T for the metronome clicks, a more complex metronome signal can be generated wherein distances between clicks are not equal but are distributed according to a more complex pattern. In particular, the pattern may correspond to a particular rhythm.

[0019] Now, according to empirical investigations of the inventors, the offsets of human drum sequences may be described by Gaussian distributed $1/f^\alpha$ noise, where f is a frequency and α is a shape parameter of the spectrum.

[0020] Figure 2 shows an example of a random signal whose power spectral density is equal to $1/f^\alpha$, wherein $\alpha = 1$, graphed double logarithmically. Within the scientific literature, this kind of noise is also referred to as 'pink noise'. The parameter α is then equivalent to the absolute value of the slope of the graph.

[0021] With regard to the invention, in particular with respect to human drumming, the parameter α may be estimated empirically by comparing the beat sequence generated by a human drum player (or several of them) with a metronome. More particularly, the temporal differences between the human and the artificial beats correspond to the off sets o_i of figure 1 and the estimation of α may be carried out by performing a linear regression on the offsets' power spectral frequency plot, wherein the frequency axis has been transformed by two logarithmic transformations for linearization.

[0022] Experiments carried out by the inventors using own recordings of the inventors as well as recordings of drummers provided by professional recording studios revealed that the exponent α appears to be widely independent of the drummer. The parameter α also clearly appears to be greater than zero (0). Also, it appears to be smaller than 2.0 in general. For drumming, it has been determined as being smaller than 1.5 in general. However, the offsets of different human drummers may differ in standard deviation and mean.

[0023] For the empirical analysis, drums have been chosen because in the analysis, the distinction between

accentuation and errors is easiest when analyzing sequences that contain time-periodic structures, such as drum sequences. However, in principle, the methods according to the invention may also be applied to other instruments played by humans. For example, for a piano player playing a song on the piano, it is expectable that after removal of accentuation, the relevant noise obeys the same $1/f^\alpha$ -law as discussed above with respect to drums.

[0024] Based on these empirically determined facts and figures, a method and a device for humanizing music, in particular drum sequences may now be described as follows.

[0025] Figure 3 shows a flowchart of a method for humanizing music sequences according to a first embodiment of the invention. The music sequence is assumed to comprise a series of sounds, which may be notes, played by an instrument such as a drum, each occurring on a distinct time t. When humanizing real audio signals, the time t may be taken as the onset of the note, which may automatically be detected by a method in the prior art (cf. e.g. Bello et al., A Tutorial on Onset Detection In Music Signals, IEEE Transactions on Speech and Audio Processing, Vol. 13, No. 5, September 2005).

[0026] In step 310, the method is initialized. In particular, the algorithm may be set to the first time to ($i = 0$).

[0027] In step 320, a random offset o_i is generated for the present sound or note at time t_i .

[0028] In step 330, the random offset o_i is added to the time t_i in order to obtain a modified time t'_i . Hereby, it is understood that the offset o_i may also be negative.

[0029] In step 340, the present sound s_i is output at the modified time t'_i . The outputting step may comprise playing the sound in an audio device. It may also comprise storing the sound on a medium, at the modified time t'_i for later playing.

[0030] In step 350, the procedure loops back to step 320 in order to repeat the procedure for the remaining sounds.

[0031] According to the invention, the random offsets are generated such that their power spectral density obeys the law

$$\frac{1}{f^\alpha}.$$

wherein $\alpha > 0$.

[0032] The parameter α may be set according to the empirical estimates obtained as described in relation to figure 2.

[0033] Figure 4 shows a block diagram of a device 400 for humanizing a music sequence according to an embodiment of the invention.

[0034] Again, it is assumed that the music sequence (S) comprises a multitude of sounds ($s_1 \dots s_n$) occurring on times ($t_1, \dots, t_n$). According to one embodiment of the

invention, the device may comprise means 410 for generating, for each time (t_i) a random offset (o_i).

[0035] The device may further comprise means 420 for adding the random offset (o_i) to the time (t_i) in order to obtain a modified time ($t_i + o_i$).

[0036] Finally, the device may also comprise means 430 for outputting a humanized music sequence (S') wherein each sound (s_i) occurs on the modified time ($t_i + o_i$).

[0037] According to the invention, the power spectral density of the random offsets has the form

$$\frac{1}{f^\alpha},$$

wherein $0 < \alpha < 2$. Generators for $1/2^\alpha$ - or 'pink' noise are commercially available.

[0038] Figure 5 shows another block diagram of a device for humanizing music sequences according to another embodiment of the invention. The device comprises a metronome 510, a noise generator 520, a module 530 for adding the random offsets to obtain a modified time sequence, a module 540 for outputting the sounds at the modified times, a module 550 for receiving an input sequence and a module 560 for analyzing the input sequence in order to automatically identify the relevant sounds.

SUMMARY

[0039] The deviation of human drum sequences from a given metronome may be well described by Gaussian distributed $1/f^\alpha$ noise, wherein the exponent α is distinct from 0. In principle, the results do also apply to other instruments played by humans. In conclusion, the method and device for humanizing musical sequence may very well be applied in the field of electronic music as well as for post processing real recordings. In other words, $1/f^\alpha$ -noise is the natural choice for humanizing a given music sequence.

Claims

1. Method for humanizing a music sequence (S), the music sequence (S) comprising a multitude of sounds ($s_1, \dots, s_n$) occurring on times ($t_1, \dots, t_n$), comprising the steps:

- generating for each time (t_i) of the times ($t_1, \dots, t_n$), a random offset (o_i), and
- adding the random offset (o_i) to the time (t_i) in order to obtain a modified time ($t_i + o_i$); and
- outputting a humanized music sequence (S') wherein each sound (s_i) occurs on the modified

time ($t_i + o_i$),

characterised in that the power spectral density of the sequence of random offsets has the form

$$\frac{1}{f^\alpha}.$$

wherein f is a frequency of the sequence of random offsets, α is a shape parameter of the spectrum and $0 < \alpha < 2$.

2. Method according to claim 1, wherein the sounds correspond to drum beats.
3. Method according to claim 1, wherein the sounds correspond to notes played by a piano.
4. Method according to claim 1, wherein the music sequence (S) is obtained from editing a human-generated music sequence.
5. Method according to claim 1, wherein the mean and/or the standard deviation of the offsets (o_i) is set according to empirical estimates.
6. Method according to claim 1, wherein outputting a humanized music sequence (S') comprises storing the music sequence (S') on a machine-readable medium.
7. Method for humanizing a music sequence (S), comprising the steps:

- obtaining (510) a regular time sequence ($t_1, \dots, t_n$);
- generating (520) a sequence of random offsets ($o_1, \dots, o_n$);
- adding (530) the sequence of random offsets ($o_1, \dots, o_n$) to the time sequence to obtain a modified time sequence ($t_1', \dots, t_n'$);
- receiving (550) an input sequence;
- analyzing (560) the input sequence in order to automatically identify sounds; and
- outputting (540) the sounds at the modified times ($t_1', \dots, t_n'$); **characterised in that** the power spectral density of the sequence of random offsets has the form

$$\frac{1}{f^\alpha}.$$

wherein f is a frequency of the sequence of random offsets, α is a shape parameter of the spectrum and $0 < \alpha < 2$.

8. Device for humanizing a music sequence (S), the music sequence (S) comprising a multitude of sounds ($s_1, \dots, s_n$) occurring on times ($t_1, \dots, t_n$), comprising:

- means for generating, for each time (t_i) of the times ($t_1, \dots, t_n$) a random offset (o_i),
- means for adding the random offset (o_i) to said each time (t_i) in order to obtain a modified time ($t_i + o_i$); and
- means for outputting a humanized music sequence (S') wherein each sound (s_i) occurs on the modified time ($t_i + o_i$),

characterised in that the power spectral density of the sequence of random offsets has the form

$$\frac{1}{f^\alpha}.$$

wherein f is a frequency of the sequence of random offsets, α is a shape parameter of the spectrum and $0 < \alpha < 2$.

Patentansprüche

1. Verfahren zum Humanisieren einer Musiksequenz (S), wobei die Musiksequenz (S) eine Vielzahl von Tönen ($s_1, \dots, s_n$) umfasst, die zu Zeiten ($t_1, \dots, t_n$) auftreten, umfassen die Schritte:

- für jede Zeit (t_i) der Zeiten (t_1 bis t_n), erzeugen eines zufälligen Abstands (o_i), und
- addieren des zufälligen Abstands (o_i) zu der Zeit (t_i), um eine modifizierte Zeit ($t_i + o_i$) zu erhalten;
- ausgeben einer humanisierten Musiksequenz (S'), in welcher jeder Ton (s_i) zu der modifizierten Zeit ($t_i + o_i$) auftritt,

dadurch gekennzeichnet, dass die spektrale Leistungsdichte der Sequenz von zufälligen Abständen die Form

$$\frac{1}{f^\alpha}$$

hat, wobei f eine Frequenz der Reihe von zufälligen Abständen, α ein Formparameter des Spektrums

und $0 < \alpha < 2$ ist.

2. Verfahren gemäß Anspruch 1, wobei die Töne Trommelschlägen entsprechen.
3. Verfahren gemäß Anspruch 1, wobei die Töne Noten sind, die mit einem Klavier gespielt werden.
4. Verfahren gemäß Anspruch 1, wobei die Musiksequenz (S) durch Editieren einer menschlich-erzeugten Musiksequenz erhalten werden.
5. Verfahren gemäß Anspruch 1, wobei das Mittel und/oder die Standardabweichung der Abstände (o_i) gemäß empirischer Schätzungen gesetzt wird.
6. Verfahren gemäß Anspruch 1, wobei das Ausgeben einer humanisierten Musiksequenz (S') ein Speichern der Musiksequenz (S') auf einem maschinenlesbaren Medium umfasst.
7. Verfahren zum Humanisieren einer Musiksequenz (S), umfassen die Schritte:

- erhalten (510) einer regelmäßigen Zeitsequenz ($t_1, \dots, t_n$);
- erzeugen (520) einer Sequenz von zufälligen Abständen ($o_1, \dots, o_n$);
- addieren (530) der Sequenz von zufälligen Abständen ($o_1, \dots, o_n$) zu der Zeitsequenz ($t_1, \dots, t_n$), um eine modifizierte Zeitsequenz ($t_1', \dots, t_n'$) zu erhalten;
- empfangen (550) einer Eingabesequenz;
- analysieren (560) der Eingabesequenz, um automatisch Töne zu identifizieren, und
- ausgeben (540) der Töne zu den modifizierten Zeiten (t_1' bis t_n');

dadurch gekennzeichnet, dass das Leistungsdichtespektrum der der Sequenz von zufälligen Abständen die Form

$$\frac{1}{f^\alpha}$$

hat, wobei f eine Frequenz der Sequenz von zufälligen Abständen ist, α ein Formparameter des Spektrums und $0 < \alpha < 2$.

8. Vorrichtung zum Humanisieren einer Musiksequenz (S), wobei die Musiksequenz (S) eine Vielzahl von Tönen ($s_1, \dots, s_n$) umfasst, die zu Zeiten ($t_1, \dots, t_n$) auftreten, umfassend:

- Mittel zum Erzeugen, für jede Zeit (t_i) der Zeiten ($t_1, \dots, t_n$) eines zufälligen Abstands (o_i),

- Mittel zum Addieren des zufälligen Abstände (o_i), zu jeder Zeit (t_i), um eine modifizierte Zeit ($t_i + o_i$) zu erhalten; und
- Mittel zum Ausgeben einer humanisierten Musiksequenz (S'), wobei jeder Ton (s_i) zu einer modifizierten Zeit ($t_i + o_i$) auftritt,

dadurch gekennzeichnet, das die spektrale Leistungsdichte der Sequenz von zufälligen Abständen die Form

$$\frac{1}{f^\alpha}$$

hat, wobei f eine Frequenz der Sequenz von zufälligen Abständen ist, α ein Formparameter des Spektrums und $0 < \alpha < 2$.

Revendications

1. Procédé d'humanisation d'une séquence musicale (S), la séquence musicale (S) comprenant une multitude de sons ($s_1, \dots, s_n$) se produisant à des moments ($t_1, \dots, t_n$), comprenant les étapes consistant à:

- pour chaque moment (t_i) des moments ($t_1, \dots, t_n$), générer un décalage aléatoire (o_i), et
- ajouter ledit décalage aléatoire (o_i), au moment (t_i) afin d'obtenir un moment modifié ($t_i + o_i$); et
- délivrer une séquence musicale humanisée (S') dans laquelle chaque son (s_i) a lieu sur ledit moment modifié ($t_i + o_i$),

caractérisé en ce que la densité de puissance spectrale de la séquence de décalages aléatoires possède la forme

$$\frac{1}{f^\alpha}.$$

où f est une fréquence de la séquence de décalages aléatoires, α est un paramètre de forme du spectre, et $0 < \alpha < 2$.

2. Procédé selon la revendication 1, dans lequel les sons correspondent à des battements de tambour.
3. Procédé selon la revendication 1, dans lequel les sons correspondent à des notes jouées par un piano.
4. Procédé selon la revendication 1, dans lequel la séquence musicale (S) est obtenue à partir de l'édition

d'une séquence musicale générée par un homme.

5. Procédé selon la revendication 1, dans lequel l'écart moyen et/ou l'écart-type des décalages (o_i) est défini selon des estimations empiriques.

6. Procédé selon la revendication 1, dans lequel la délivrance d'une séquence musicale humanisée (S') comprend le stockage de la séquence musicale (S') sur un support lisible par une machine.

7. Procédé d'humanisation d'une séquence musicale (S'), comprenant les étapes consistant à:

- obtenir (510) une séquence temporelle régulière ($t_1, \dots, t_n$);
- générer (520) une séquence de décalages aléatoires ($o_1, \dots, o_n$);
- ajouter (530) la séquence de décalages aléatoires ($o_1, \dots, o_n$) à la séquence temporelle ($t_1, \dots, t_n$) afin d'obtenir une séquence temporelle modifiée ($t'_1, \dots, t'_n$);
- recevoir (550) une séquence d'entrée;
- analyser (560) la séquence d'entrée afin d'identifier automatiquement les sons; et
- délivrer (540) les sons aux moments modifiés ($t'_1, \dots, t'_n$);

caractérisé en ce que la densité de puissance spectrale de la séquence de décalages aléatoires possède la forme

$$\frac{1}{f^\alpha}.$$

où f est une fréquence des décalages aléatoires, α est un paramètre de forme du spectre, et $0 < \alpha < 2$.

8. Dispositif d'humanisation d'une séquence musicale (S), la séquence musicale (S) comprenant une multitude de sons ($s_1, \dots, s_n$) se produisant à des moments ($t_1, \dots, t_n$), comprenant:

- un moyen de génération, pour chaque moment (t_i) des moments ($t_1, \dots, t_n$), d'un décalage aléatoire (o_i),
- un moyen d'ajout du décalage aléatoire (o_i) à chacun desdits moments (t_i) afin d'obtenir un moment modifié ($t_i + o_i$); et
- un moyen de délivrance d'une séquence musicale humanisée (S') dans laquelle chaque son (s_i) a lieu audit moment modifié ($t_i + o_i$),

caractérisé en ce que la densité de puissance spectrale de la séquence de décalages aléatoires possède la forme

$$\frac{1}{f^\alpha}.$$

où f est une fréquence de la séquence de décalages aléatoires, α est un paramètre de forme du spectre, et $0 < \alpha < 2$.

5

10

15

20

25

30

35

40

45

50

55

Fig. 1

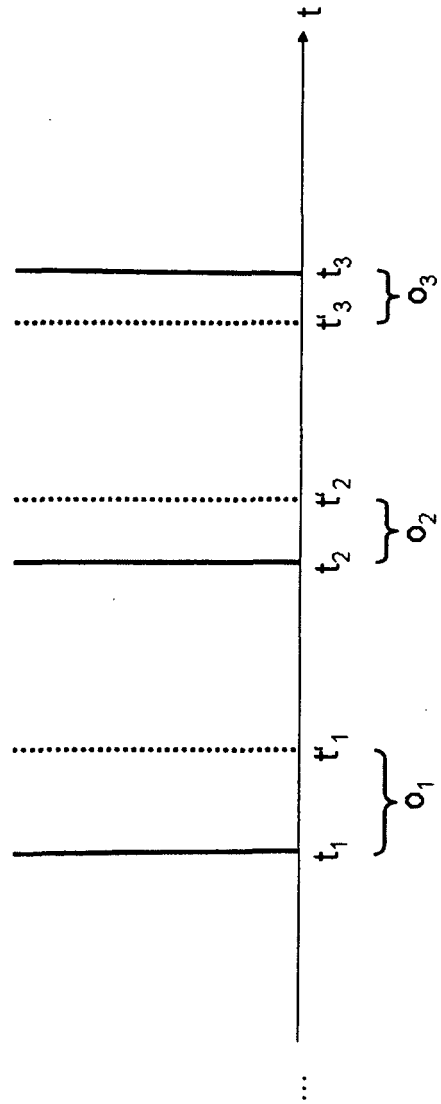


Fig. 2

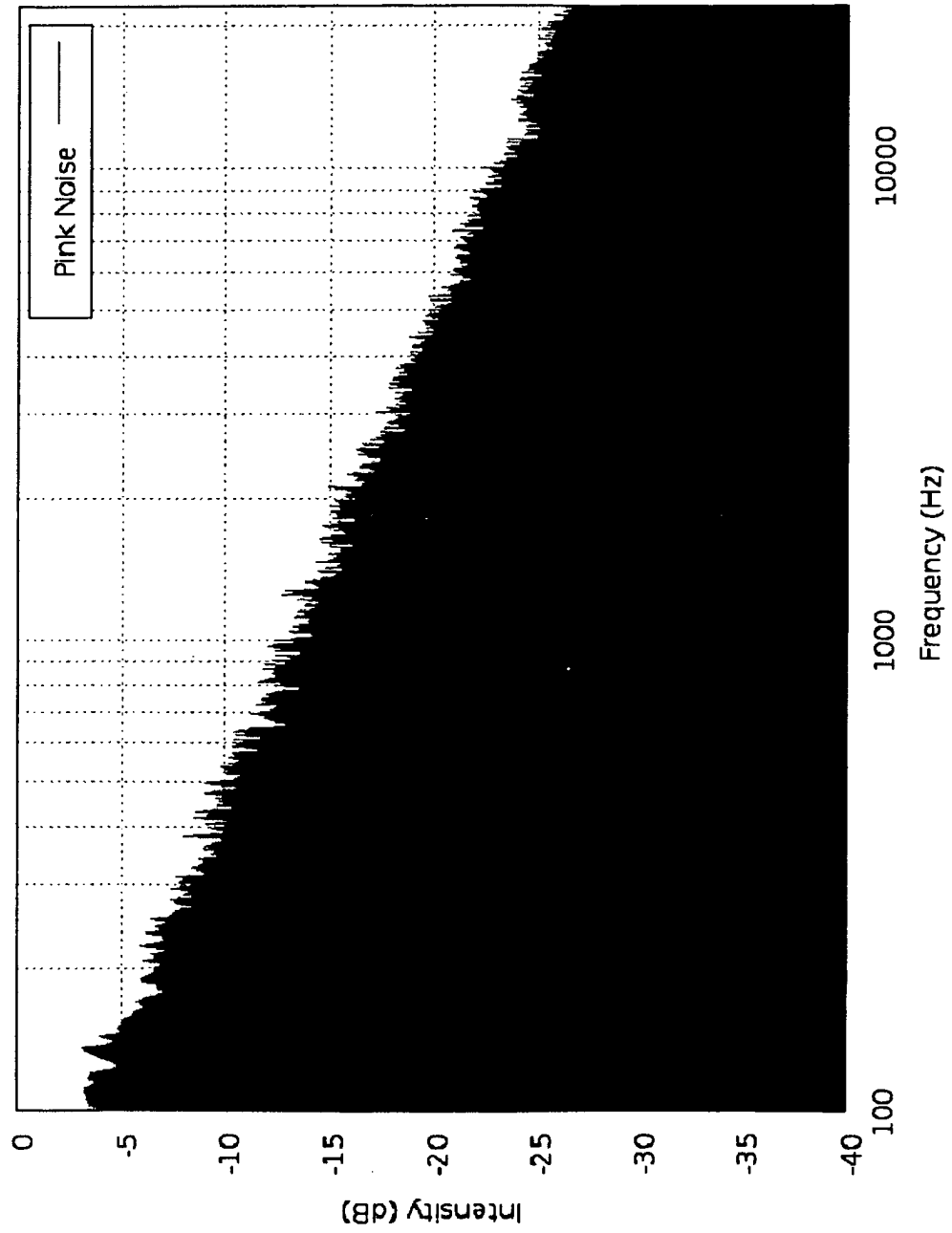


Fig. 3

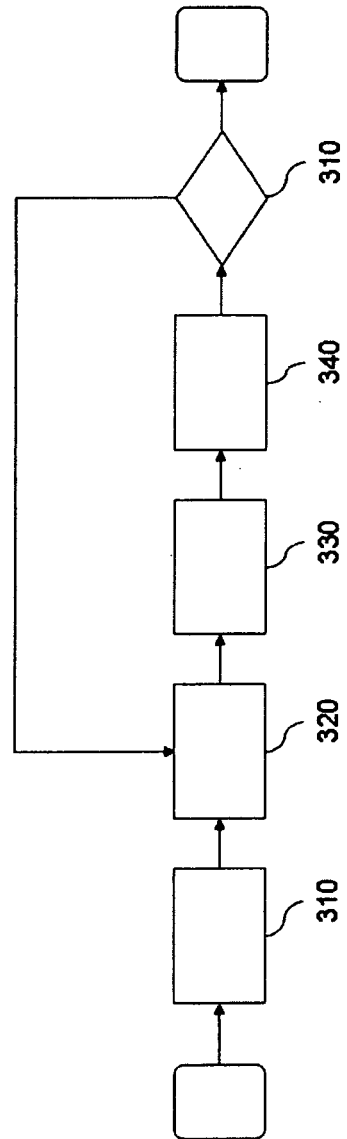


Fig. 4

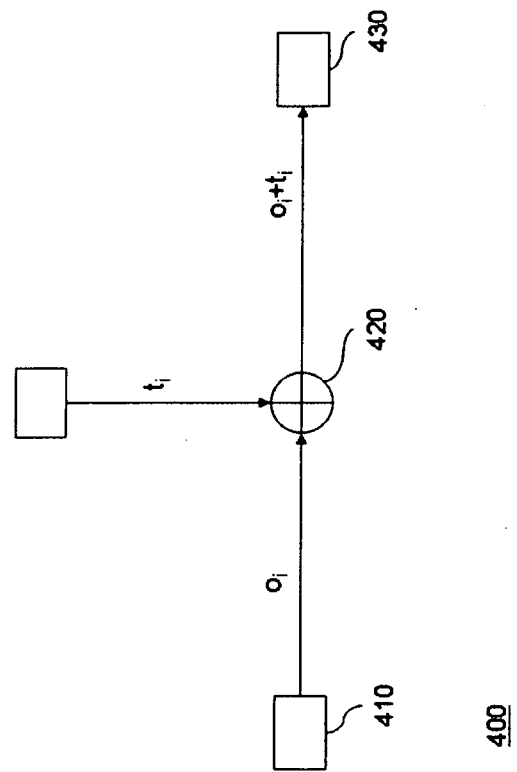
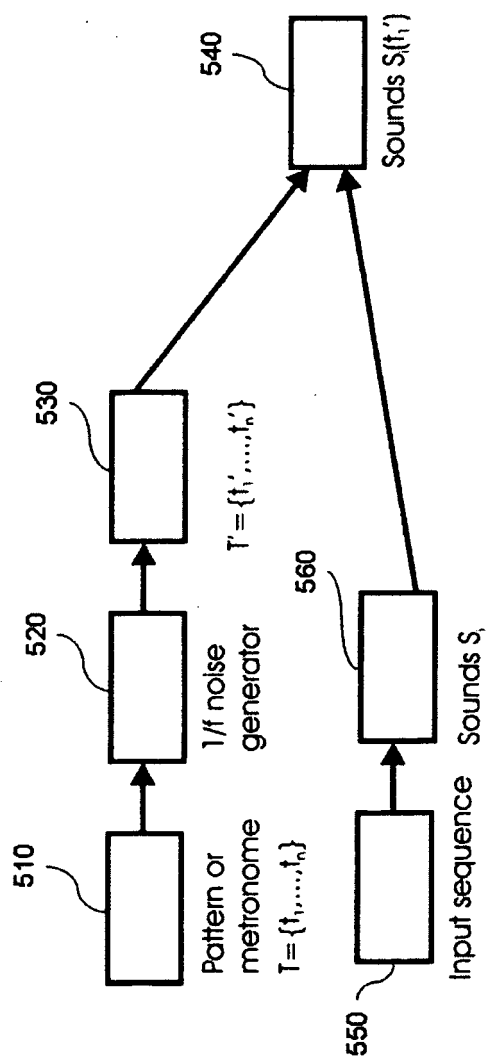


Fig. 5



500

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 3974729 A [0006]
- US 6066793 A [0006]
- US 6506969 B [0007]
- GB 2418054 A [0008]

Non-patent literature cited in the description

- **BELLO et al.** A Tutorial on Onset Detection In Music Signals. *IEEE Transactions on Speech and Audio Processing*, September 2005, vol. 13 (5 [0025]