



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
08.10.2014 Bulletin 2014/41

(51) Int Cl.:
G10L 19/018 (2013.01)

(21) Application number: **13162596.4**

(22) Date of filing: **05.04.2013**

(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

- **Barni, Mauro**
I-59100 Prato (IT)
- **Ferrari, Gianluigi**
I-43124 Parma (IT)
- **Arconte, Davide**
I-20124 Milano (IT)

(71) Applicant: **Movym S.r.l.**
20124 Milano (IT)

(74) Representative: **Giannessi, Simona**
Porta, Checcacci & Associati S.p.A.
Via Trebbia, 20
20135 Milano (IT)

(72) Inventors:
 • **Abrardo, Andrea**
I-50018 Scandicci (FI) (IT)

(54) **Method and system of audio signal watermarking**

(57) Method and system of watermarking an audio signal wherein:
 - the audio signal is assigned to a class, among a plurality of classes, depending on the semantic content of the audio signal, the plurality of classes being associated with a corresponding plurality of watermark profiles;

- the watermark profile associated with the class assigned to the audio signal is obtained;
 - a watermark is embedded into the audio signal by using the obtained watermark profile so as to provide a watermarked audio signal.

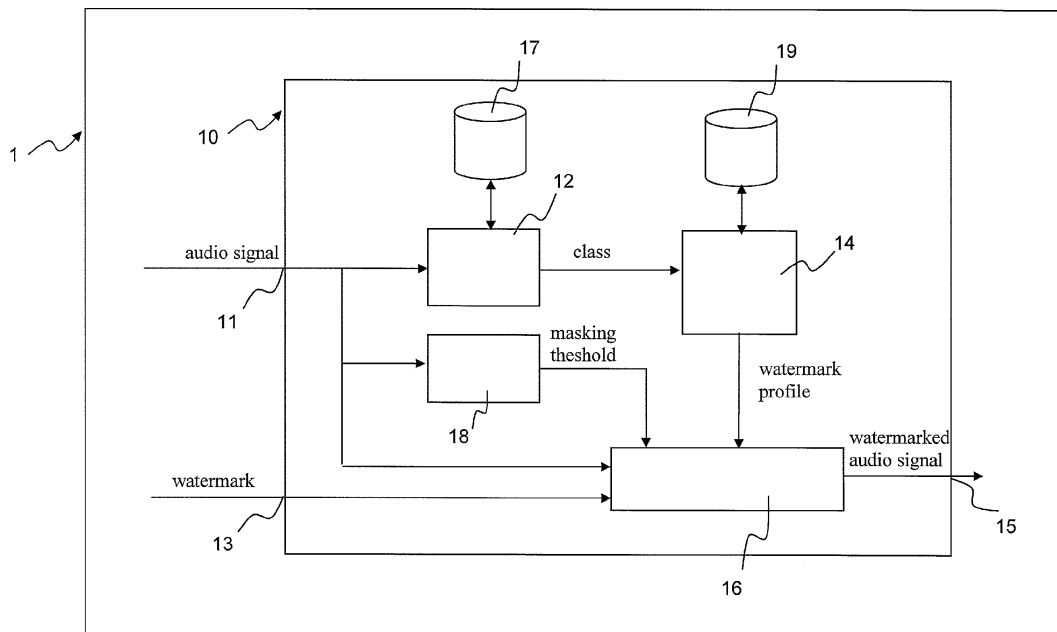


Fig. 1

Description

[0001] The present invention relates to a method and system of audio signal watermarking.

[0002] Audio signal watermarking is a process for embedding information data (watermark) into an audio signal without affecting the perceptual quality of the host signal itself.

[0003] The watermark should be imperceptible or nearly imperceptible to Human Auditory System (HAS). However, the watermark should be detectable through an automated detection process.

[0004] Watermarking techniques are known in the art.

[0005] For example, EP 1 594 122 discloses a watermarking method and apparatus employing spread spectrum technology and psycho-acoustic model. According to spread spectrum technology, a small baseband signal bandwidth is spread over a larger bandwidth by injecting or adding a higher frequency signal, or spreading function. Thereby the energy used for transmitting the signal is spread over a wider bandwidth, and appears as noise. According to psycho-acoustic model, based on psychoacoustical properties of the HAS, the watermark signal is shaped to reduce its magnitude so that it has a level that is below a masking threshold of the host audio/video signal. In the method and apparatus disclosed by EP 1 594 122, at encoder side a spreading function is modulated by watermark data bits for providing a watermark signal; the current masking level of the audio/video signal is determined and a corresponding psycho-acoustic shaping of the watermark signal is performed; the psycho-acoustically shaped watermark signal is additionally shaped in order to reduce on average the magnitude of the watermark signal, whereby for each spectral line the phase of the values of the audio/video signal into which the psycho-acoustically and additionally shaped watermark signal is embedded is kept unchanged by the additional shaping; the psycho-acoustically and additionally shaped watermark signal is embedded into the audio/video signal.

[0006] The applicant observes that a watermarking technique should achieve a trade-off between three basic features: imperceptibility, robustness and payload, which are strictly linked to each other by inverse relationships. Depending upon the purpose of using watermarking, a watermarking technique should find a correct balance between the need to keep the watermark imperceptible; to make the watermark robust against attacks and manipulations of the host signal (e.g., noise distortion, A/D or D/A conversion, lossy coding, resizing, filtering, lossy compression); and to achieve the highest possible payload.

[0007] The applicant notes that psycho-acoustic model enables to determine the maximum distortion, i.e. the maximum watermark signal energy that can be introduced into a host signal without being perceptible by human senses. However, this model does not provide any information about robustness and payload and about op-

timization of trade-off among imperceptibility, robustness and payload.

[0008] It is thus an object of the invention to provide alternative method and system of audio signal watermarking.

[0009] It is a further object of the invention to provide improved method and system of audio signal watermarking with high performances in terms of trade-off among imperceptibility, robustness and payload.

[0010] The Applicant found that the above objects are achieved by a method and system of audio signal watermarking wherein audio signals are classified based on their semantic content and watermarks are embedded into the audio signals by using watermark profiles selected on the basis of the classes assigned to the audio signals.

[0011] Indeed, as described in more detail below, the Applicant found that, given an audio signal, the trade-off among watermark imperceptibility, robustness and payload can be optimized by fitting the watermark profile depending on the semantic content of the audio signal.

[0012] In the present disclosure, the expression "semantic content" in relation to an audio signal refers to the audio type contained in the audio signal. The semantic content of an audio signal can be, for example, speech (e.g. talks from movies, from TV or radio programs, from TV or radio advertisements, from TV or radio talk shows, and similar) or music. In case of music, the semantic content can be, for example, a musical genre (e.g., rock, classic, jazz, blues, instrumental, singing and similar). In case of speech, the semantic content can be, for example, a tone of voice, conversation, single person speaking, whisper, aloud quarrel, and similar.

[0013] In the present disclosure, the expression "watermark profile" is used to indicate a set of parameters used for embedding the watermark into the audio signal according to a predetermined watermarking technique.

[0014] In a first aspect, the present disclosure relates to a method of watermarking an audio signal comprising:

- assigning the audio signal to a class, among a plurality of classes, depending on the semantic content of the audio signal, the plurality of classes being associated with a corresponding plurality of watermark profiles;
- obtaining the watermark profile associated with the class assigned to the audio signal;
- embedding a watermark into the audio signal by using the obtained watermark profile so as to provide a watermarked audio signal.

[0015] In a second aspect, the present disclosure relates to a system of watermarking an audio signal comprising an encoding device comprising:

- a classification unit configured to assign the audio

signal to a class, among a plurality of classes, depending on the semantic content of the audio signal, the plurality of classes being associated with a corresponding plurality of watermark profiles;

- a watermark profile unit configured to obtain the watermark profile associated with the class assigned to the audio signal;
- an embedding unit configured to embed a watermark into the audio signal by using the watermark profile obtained by watermark profile unit so as to provide a watermarked audio signal.

[0016] The method and system of the present disclosure may have at least one of the following preferred features.

[0017] Advantageously, each watermark profile is associated with a corresponding class so that trade-off among watermark imperceptibility, robustness and payload is optimized for said class, depending on the watermark application. For example, depending on the watermark application, one, two or all the features among imperceptibility, robustness and payload could be optimized for each class, by keeping unchanged the other feature(s), if any, among the classes. For example, in noisy applications, robustness could be maximized for each class, by keeping the same payload and imperceptibility level among the classes. On the other hand, in low-noise application and/or when many data need to be contained in the watermark, payload could be maximized for each class, by keeping the same robustness and imperceptibility level among the classes. Otherwise, in low-noise application, imperceptibility could be maximized for each class, by keeping the same payload and robustness level among the classes.

[0018] The plurality of watermark profiles can relate to a single watermarking technique or to a plurality of watermarking techniques.

[0019] In case of single watermarking technique, the watermark profiles differ from each other for the value taken by at least one parameter of said set of parameters.

[0020] In case of a plurality of watermarking techniques, the watermark profiles differ from each other for at least one of the parameters and/or for the values taken by at least one of the common parameters.

[0021] The watermarking technique(s) can be selected, for example, from the group comprising: spread spectrum watermarking technique, echo hiding watermarking technique, phase coding technique, informed watermarking schemes like QIM (Quantization Index Modulation) and Spread Transform Dither Modulation (STDM).

[0022] In a preferred embodiment, the method is a computer-implemented method.

[0023] In an embodiment, the embedding unit can comprise a plurality of embedding sub-units. Each embedding sub-unit can be configured to embed the watermark into the audio signal by using one watermark profile

of said plurality of watermarking profiles or one watermarking technique of said plurality of watermarking techniques.

[0024] At least one parameter of said set of parameters defining the watermark profile may be selected, for example from the group comprising: watermark bit rate; frequency range hosting the watermark; Document to Watermark Ratio (DWR); watermark frame length; masking threshold modulation factor F , intended as a quantity by which the masking threshold of the audio signal -computed according to a psychoacoustic model- is multiplied to vary its amplitude with respect to the computed value; channel coding scheme (which may also include error detection techniques such as, for example, Cyclic Redundancy Check); number, amplitude, offset and decay rate of echo pulses (in case of echo hiding watermarking technique); spreading factor, intended as number of audio signal frequency or phase samples needed to insert one bit of watermark (in case of spread spectrum watermarking technique with, respectively, frequency or phase modulation).

[0025] Preferably, the plurality of classes is associated with the corresponding plurality of watermark profiles in a database. The database can be internal or external to the encoding device.

[0026] In a preferred embodiment, the expression watermarking refers to digital watermarking.

[0027] Digital watermarking relates to a computer-implemented watermarking process.

[0028] In a preferred embodiment, a masking threshold of the audio signal according to a psychoacoustic model is computed. Preferably, before computing the masking threshold, the audio signal is split into time windows and a masking threshold is computed for each time window of the audio signal. The masking threshold can be computed after, before or at the same time of the audio signal classification.

[0029] The psychoacoustic model can be a psychoacoustic model known in the art.

[0030] Preferably, the psychoacoustic model is adapted to calculate the masking threshold in time and/or frequency domain and is based on one of the following analysis: block based FFT (Fast Fourier Transform), block based DCT (Discrete Cosine Transform), block based MDCT (Modified Discrete Cosine Transform), block based MCLT (Modified Complex Lapped Transform), block based STFT (Short-Time Fourier Transform), sub-band or wavelet packet analysis.

[0031] Preferably, the encoding device comprises a masking unit configured to perform the masking threshold computation and, optionally, the time windows splitting.

[0032] Preferably, embedding the watermark into the audio signal comprises the step of shaping the energy of the watermark according to the computed masking threshold.

[0033] This advantageously enables to guarantee watermark imperceptibility to human auditory system.

[0034] Preferably, the watermark is shaped to reduce its energy below the computed masking threshold of the audio signal. When the set of parameters defining the obtained watermark profile comprises the masking threshold modulation factor F, the watermark is preferably shaped to reduce its energy below the computed masking threshold, multiplied by the masking threshold modulation factor F.

[0035] Preferably, the watermark is shaped by the embedding unit.

[0036] Preferably, the masking threshold modulation factor F is at least equal to 0.5. More preferably, the masking threshold modulation factor F is at least equal to 0.7, even more preferably at least equal to 0.8, even more preferably at least equal to 0.9.

[0037] Preferably, the masking threshold modulation factor F is not higher than 1.5. More preferably, the masking threshold modulation factor F is not higher than 1.3, even more preferably not higher than 1.2, even more preferably not higher than 1.1.

[0038] Assigning the audio signal to a class, according to the semantic content of the audio signal, is preferably performed based upon analysis of at least one audio signal feature.

[0039] In the present disclosure, the audio signal feature is related to the semantic content of the audio signal.

[0040] In the present disclosure, the audio signal feature is preferably related to time, frequency, energy or cepstrum domain.

[0041] For example, the at least one audio signal feature can be selected from the group comprising: loudness, brightness, beats per minute (BPM) bandwidth, pitch, odd-to-even harmonic energy ratio, spectral energy bands (e.g. spectrum sparsity), spectral and tonal complexities, spectral roll-off point (intended as any percentile of the power spectral distribution), spectral centroid (defined as the center of gravity of the magnitude spectrum), spectral "flux" (intended as squared difference between the normalized magnitudes of successive spectral distributions), time domain Zero-Crossing Rate, Cepstrum Resynthesis Residual Magnitude, Mel-Frequency Cepstral Coefficients, band periodicity (defined as the periodicity of a sub-band and derived by sub-band correlation analysis).

[0042] The analysis of at least one audio signal feature preferably comprises checking if the at least one audio signal feature meets one or more predetermined constraints. For example, the value of the at least one audio signal feature can be compared to one or more predetermined thresholds or one or more predetermined ranges of values. Each class is advantageously defined by predetermined constraints (e.g. set of values) to be met by the at least one audio signal feature.

[0043] Preferably, before assigning the audio signal to a class, the audio signal is split into sub-signals of shorter duration and a class is assigned to each sub-signal independently from the other sub-signals.

[0044] Suitably, the duration of the sub-signals is long-

er than the duration of the time windows in which the audio signal is split for performing masking threshold computation.

[0045] Preferably, the method of audio signal watermarking comprises a decoding process comprising extraction of the watermark from the watermarked audio signal. Preferably, the watermark is extracted by using the same watermark profile used for embedding the watermark into the audio signal.

[0046] Preferably, the system also comprises a decoding device configured to extract the watermark from the watermarked audio signal.

[0047] In an embodiment of the decoding process, the watermarked audio signal is assigned to a class, among the plurality of classes, depending on the semantic content of the watermarked audio signal, the plurality of classes being associated with the corresponding plurality of watermark profiles. Preferably, the class is assigned by a classification unit of the decoding device. According to this embodiment, the watermark profile associated with the class assigned to the audio signal is obtained and used for extracting the watermark from the watermarked audio signal. Preferably, the watermark is extracted from the watermarked audio signal by an extraction unit of the decoding device.

[0048] According to another embodiment of the decoding process, said plurality of watermark profiles are tried in sequence for extracting the watermark till the watermark is successfully extracted from the watermarked audio signal. In this case, the decoding device can comprise a single extraction unit for trying in sequence the plurality of watermark profiles. In alternative, the extraction unit can comprise a plurality of sub-extraction units, one for each watermark profile or for each watermarking technique, for trying the plurality of watermark profiles at least partly in parallel. In this embodiment, audio signal classification is not necessary at the decoding side.

[0049] According to a further embodiment, a second watermark is embedded into the audio signal, comprising the class assigned to the audio signal, by using a predefined watermark profile, common to all audio signals independently from their class. The second watermark can be embedded into the watermarked audio signal, already watermarked with the first watermark. In alternative, the first and the second watermarks can be embedded into different sub-bands of the audio signal. Watermark extraction can then be performed by first extracting the second watermark from the watermarked audio signal (by using the common watermark profile) so as to retrieve the class of the audio signal, and then by obtaining the watermark profile associated with the retrieved class and extracting the watermark from the watermarked audio signal with the obtained watermark profile. In this embodiment, audio signal classification is not necessary at the decoding side.

[0050] Further characteristics and advantages of the present invention will become clearer from the following detailed description of some preferred embodiments

thereof, made as an example and not for limiting purposes with reference to the attached drawings. In such drawings,

- figure 1 schematically shows a system of audio signal watermarking according to an embodiment of the invention;
- figure 2 schematically shows the energy spectrum of four audio signals having a different semantic content;
- figure 3 schematically shows a system of audio signal watermarking according to another embodiment of the invention;
- figure 4 schematically shows a first embodiment of a decoding device of the system of figure 3;
- figure 5 schematically shows a second embodiment of the decoding device of the system of figure 3;
- figure 6 schematically shows an exemplary implementation of the system of figure 3.

[0051] Figure 1 discloses a system 1 of audio signal watermarking according to an embodiment of the invention.

[0052] The system 1 comprises an encoding device 10 comprising an input 11 for an audio signal, an input 13 for a watermark and an output 15 for a watermarked audio signal.

[0053] The encoding device 10 comprises a classification unit 12, a watermark profile unit 14, a masking unit 18 and an embedding unit 16.

[0054] The classification unit 12, watermark profile unit 14, masking unit 18 and embedding unit 16 comprise hardware and/or software and/or firmware configured to implement the method of the present disclosure.

[0055] The classification unit 12 is configured to assign the audio signal to a class depending on the semantic content of the audio signal.

[0056] The classification unit 12 is configured to analyse at least one audio signal feature related to the semantic content of the audio signal, to compare the at least one audio signal feature with one or more constraints and to assign the audio signal a class, selected among a predetermined plurality of classes, depending on the result of the comparison.

[0057] For example, the at least one audio signal feature can be selected from the group comprising: loudness, brightness, beats per minute (BPM) bandwidth, pitch, odd-to-even harmonic energy ratio, spectral energy bands (e.g. spectrum sparsity), spectral and tonal complexities, spectral roll-off point (intended as any percentile of the power spectral distribution), spectral centroid (defined as the center of gravity of the magnitude spectrum), spectral "flux" (intended as squared differ-

ence between the normalized magnitudes of successive spectral distributions), time domain Zero-Crossing Rate, Cepstrum Resynthesis Residual Magnitude, Mel-Frequency Cepstral Coefficients, band periodicity (defined as the periodicity of a sub-band and derived by sub-band correlation analysis).

[0058] For example, the at least one audio signal feature can be compared with one or more predetermined thresholds or one or more predetermined ranges of values and each class can be defined by a predetermined set of values that can be taken by the at least one audio signal feature.

[0059] The plurality of classes can be stored in a suitable class database 17 internal (as shown in figure 1) or external to the encoding device 10.

[0060] In a preferred embodiment (not shown), before assigning the audio signal to a class, the encoding device 10 is configured to split the audio signal into sub-signals of shorter duration (e.g. from few tenths to few tens of seconds) and the classification unit 12 is configured to classify each sub-signal independently from the other sub-signals.

[0061] In a preferred embodiment, the classification unit 12 is configured to classify the audio signals (or sub-signals) by analysing the spectrum sparsity of their energy spectrum.

[0062] The spectrum sparsity is an audio signal feature indicative of the energy concentration in a sub-band compared to the energy in the whole audio signal (or sub-signal) band.

[0063] The energy spectrum of the audio signal (or sub-signal) is considered sparse (or colored) if most part of its energy is concentrated in a small spectrum sub-band, otherwise it is considered non-sparse (or noise-like).

[0064] For example, figure 2 shows the energy spectrum of four audio signals having a different semantic content: speech, rock, jazz, piano solo.

[0065] For example, in case of figure 2, three different classes can be defined by analyzing the spectrum sparsity feature and, in the example, by comparing the fraction of signal energy (normalized to the total energy) contained in the 0-1000 Hz sub-band with two threshold levels S_L and S_H . If said fraction of energy in the 0-1000 Hz sub-band is lower than S_L , the audio signal (or sub-signal) can be classified into a "low sparse" class; if said fraction of energy in the 0-1000 Hz sub-band is between S_L and S_H , the audio signal (or sub-signal) can be classified into a "medium sparse" class; if it is higher than S_H , the audio signal can be classified into a "high sparse" class.

[0066] In the example of figure 2, by setting $S_L = 0.85$ and $S_H = 0.90$, talk and rock signals are classified as "low sparse" signals, jazz signal is classified as "medium sparse" signal, and piano solo signal is classified as "high sparse" signal.

[0067] It is noted that most audio signals have their energy concentrated in the 0-1000 Hz sub-band. Even slight differences (e.g. of about 0.05) between threshold

levels S_L and S_H can thus be significant in such sub-band.

[0068] The plurality of classes used by the classification unit 12 are associated with a corresponding plurality of watermark profiles in a suitable watermark profile database 19 internal (as shown in figure 1) or external to the encoding device 10.

[0069] It is noted that, even if database 17 and database 19 are shown in the figures as two distinct entities, they can also be implemented into a single database.

[0070] Each watermark profile is defined by a set of parameters used for embedding the watermark into the audio signal according to a predetermined watermarking technique.

[0071] The watermarking technique can be a technique known in the art as, for example, a spread spectrum watermarking technique (e.g. wherein the watermark is spread over many frequency bins so that the energy in one bin is very small and undetectable), an echo hiding watermarking technique (e.g. wherein the watermark is embedded into an audio signal by introducing one or more echoes that are offset in time from the audio signal by an offset value associated with the data value of the bit), a phase coding technique (e.g. wherein phase differences between selected frequency component portions of an audio signal are modified to embed the watermark in the audio signal) or any other watermarking technique known in the art.

[0072] The set of parameters can comprise at least one parameter selected from the group comprising: watermark bit rate; frequency range hosting the watermark; Document to Watermark Ratio (DWR); watermark frame length; masking threshold modulation factor F ; channel coding scheme; number, amplitude, offset and decay rate of echo pulses; spreading factor.

[0073] The plurality of watermark profiles associated with the plurality of classes can relate to a single watermarking technique or to a plurality of watermarking techniques.

[0074] In case of single watermarking technique, the watermark profiles are all defined by the same set of parameters and differ from each other for the values taken by at least one of the parameters.

[0075] In case of a plurality of watermarking techniques, the watermark profiles relating to different watermarking techniques are defined by different sets of parameters. The watermark profiles can thus differ from each other for at least one parameter and/or for at least one value taken by a common parameter.

[0076] According to the present disclosure, within the watermark profile database 19, each class is associated with a corresponding watermark profile that enables to optimize trade-off among watermark imperceptibility, robustness and payload for each class, depending on the watermark application.

[0077] In fact, the applicant observed that it is possible to obtain different optimized trade-offs for the audio signals, depending on the semantic content of each audio signal.

[0078] For example, as visible in figure 2, talk and rock signals, classified as "low sparse" signals, allows to introduce a higher level of distortion than jazz signals, classified as "medium sparse" signals, and than piano solo signals, classified as "high sparse" signals.

[0079] According to the invention, the level of distortion which is actually "available" for each audio signal class is advantageously exploited in order to optimize trade-off among watermark imperceptibility, robustness and payload, depending on the watermark application. For example, the higher level of distortion available for the "low sparse" and "medium sparse" signals, compared with the "high sparse" signals, could be exploited to maximize, for each class, one or two features among imperceptibility, robustness and payload, by keeping unchanged the other feature(s) among the classes.

[0080] For example, when the audio signal is intended to be transmitted in a low noise channel and/or played in a low-noise ambient (e.g. domestic ambient), payload could be maximized for each class, by keeping the same robustness and imperceptibility levels among the classes. On the other hand, when the audio signal is intended to be transmitted in a high noise channel and/or played in a high-noise environment (e.g. public ambient as a train station or airport), robustness could be maximized for each class, by keeping the same payload and imperceptibility level among the classes. Otherwise, when the audio signal is intended to be played in a low-noise ambient, imperceptibility could be maximized for each class, by keeping the same payload and robustness level among the classes.

[0081] Within the set of parameters defining a watermark profile, payload could be optimized by acting on the watermark bit rate; imperceptibility could be optimized by acting on the masking threshold modulation factor F ; and robustness could be optimized by acting on at least one of: frequency range hosting the watermark; Document to Watermark Ratio; watermark frame length; channel coding scheme; number, amplitude, offset and decay rate of echo pulses; and spreading factor.

[0082] For example, in the case of figure 2, supposing to keep the same robustness and imperceptibility levels among classes, the level of distortion actually "available" for each audio signal class could be exploited in order to maximize the payload feature for each class, by associating a watermark profile with a higher bit rate with the low sparse class, a watermark profile with an intermediate bit rate with the medium sparse class and a watermark profile with a lower bit rate with the high sparse class. On the other hand, supposing to keep the same payload and imperceptibility level among classes, the robustness feature could be maximized for each class, by associating a more robust watermark profile with the low sparse class, an intermediate robust watermark profile with the medium sparse class and a lower robust watermark profile with the high sparse class. Otherwise, supposing to keep the same payload and robustness level among classes, the imperceptibility feature could be

maximized for each class, by associating, for example, a watermark profile having a masking threshold modulation factor $F \geq 1$ with the low sparse class, a watermark profile having a masking threshold modulation factor $F = 1$ with the medium sparse class and a watermark profile having a masking threshold modulation factor $F \leq 1$ with the high sparse class.

[0083] As to the masking threshold modulation factor F , it is also observed that it could be set to different values, depending on the frequency ranges of the audio signal.

[0084] As also explained in more detail below, in case of $F = 1$, the watermark is shaped by the embedding unit 16 according to a masking threshold as computed by the masking unit 18, on the basis of a psycho-acoustic model. In case of $F > 1$, the watermark is shaped according to a higher masking threshold, whereby the imperceptibility level of the watermark is decreased with respect to the level set according to the psycho-acoustic model. In case of $F < 1$, the watermark is shaped according to a lower masking threshold, whereby the imperceptibility level of the watermark is increased with respect to the level set according to the psycho-acoustic model.

[0085] This is advantageous because the psychoacoustic model, as such, produces a representation of sound perception based on average human auditory system, without taking into account high or low level psychoacoustic effects. Indeed, the masking threshold computed according to the psychoacoustic model can be too strict in some situations (e.g. in case of rock music, noisy signal, conversation or aloud quarrel) or too light in other situations (e.g. in case of classic or instrumental music and expert hearer).

[0086] Accordingly, the masking threshold modulation factor F enables to vary the amplitude of the masking threshold, as computed according to the psycho-acoustic model, depending on the semantic content of the audio signal. In this way, the imperceptibility level of the watermark can be finely tuned, with respect to the level set according to the psycho-acoustic model, depending on the semantic content of the audio signal and on the watermark application.

[0087] Once the audio signal is assigned to a class by the classification unit 12, the watermark profile unit 14 is configured to retrieve from the watermark profile database 19 the watermark profile associated with said class and to provide it to the embedding unit 16.

[0088] The masking unit 18 is configured to compute a masking threshold of the audio signal according to a psycho-acoustic model and to provide it to the embedding unit 16.

[0089] The psychoacoustic model can be any psychoacoustic model known in the art.

[0090] Preferably, the psychoacoustic model calculates the masking threshold in time and/or frequency domain and is based on one of the following analysis: block based FFT (Fast Fourier Transform), block based DCT (Discrete Cosine Transform), block based MDCT (Modified Discrete Cosine Transform), block based MCLT

(Modified Complex Lapped Transform), block based STFT (Short-Time Fourier Transform), sub-band or wavelet packet analysis.

[0091] Preferably, the masking unit 18 is configured to split the audio signal into suitable time windows (e.g. of about few ms) and to compute a masking threshold for each time window.

[0092] In the embodiment shown, the masking threshold computation is performed in parallel to audio signal classification.

[0093] The embedding unit 16 is configured to embed the watermark into the audio signal by using the watermark profile obtained by the watermark profile unit 14 so as to provide a watermarked audio signal.

[0094] The embedding unit 16 is also configured to shape the energy of the watermark according to the masking threshold computed by the masking unit 18.

[0095] When the set of parameters defining the watermark profile obtained by the watermark profile unit 14 comprises the masking threshold modulation factor F , the embedding unit 16 is also preferably configured to shape the watermark so as to reduce its energy below the masking threshold computed by the masking unit 18, multiplied by the masking threshold modulation factor F .

[0096] In an embodiment (not shown) the embedding unit 16 can comprise a plurality of embedding sub-units, one for each different watermark profile of said plurality of watermarking profiles or one for each of the watermarking techniques to which the plurality of watermarking profiles relates.

[0097] Figure 3 shows an embodiment wherein the system 1 comprises the encoding device 10, a communication network 30 and a decoding device 20.

[0098] As far as the encoding device 10 is concerned reference is made to what disclosed above.

[0099] The communication network 30 and the decoding device 20 comprise hardware and/or software and/or firmware configured to implement the method of the present disclosure.

[0100] The communication network 30 can be any type of communication network adapted to transmit the watermarked audio signal.

[0101] The decoding device 20 is configured to receive the watermarked audio signal and to extract the watermark from it.

[0102] Preferably, the watermark is extracted by using the same watermark profile used for embedding the watermark into the audio signal. In view of this, the decoding device 20 needs to know the watermark profile used for embedding the watermark.

[0103] Figure 4 shows a first embodiment of the decoding device 20 comprising a classification unit 22, a watermark profile unit 24, an extraction unit 26, a class database 27 and a watermark profile database 29.

[0104] The classification unit 22 is configured to assign the watermarked audio signal a class depending upon the semantic content of the audio signal, in the same way to what disclosed above with reference to classification

unit 12 of encoding device 10. The class assigned to the watermarked audio signal will thus be the same as that assigned in the encoding device 10.

[0105] The class database 27 (which could also be external to the decoding device 20) stores the plurality of classes in which the audio signal can be classified.

[0106] The watermark profile database 29 (which could also be external to the decoding device 20) stores an association between the plurality of classes and the corresponding plurality of watermark profiles.

[0107] It is noted that, even if database 27 and database 29 are shown in the figures as two distinct entities, they can also be implemented into a single database.

[0108] Once the audio signal is assigned to a class by the classification unit 22, the watermark profile unit 24 is configured to retrieve from the watermark profile database 29 the watermark profile associated with said class and to provide it to the extraction unit 26.

[0109] The association in the watermark profile database 29 is the same as that in the watermark profile database 19 of the encoding device 10. The watermark profile retrieved by the watermark profile unit 24 will thus be the same as that used in the encoding device 10.

[0110] The extraction unit 26 is configured to use the watermark profile retrieved by watermark profile unit 24 for extracting the watermark from the watermarked audio signal.

[0111] Figure 5 shows a second embodiment of the decoding device 20 comprising an extraction unit 26. In this embodiment, watermarked audio signal classification is not performed in decoding device 20.

[0112] According to a first implementation of this embodiment, extraction unit 26 is configured to try in sequence the plurality of watermark profiles for extracting the watermark till the watermark is successfully extracted from the watermarked audio signal. The extraction unit 26 can comprise a plurality of extraction sub-units (not shown), one for each watermark profile or for each watermarking technique, for trying the plurality of watermark profiles at least partly (or wholly) in parallel.

[0113] According to a second implementation, the class of the audio signal can be inserted into the audio signal by the embedding unit 16 of the encoding device 10 by embedding a second watermark (containing said class) into the audio signal with a predefined watermark profile, common to all audio signals independently from their class. The second watermark can be embedded into the already watermarked audio signal. In alternative, the two watermarks can be embedded into different sub-bands of the audio signal.

[0114] In this second variant, the extraction unit 16 is preferably configured to first use the predefined common watermark profile to extract the second watermark from the watermarked audio signal thereby retrieving the class of the audio signal. Then, the extraction unit 16 is configured to obtain (e.g., from a watermark profile database - not shown in figure 5- similar to watermark profile database 29) the watermark profile associated with the re-

trieved class and to extract the watermark from the watermarked audio signal with the obtained watermark profile.

[0115] Figure 6 shows an exemplary implementation of the system 1 of audio signal watermarking.

[0116] According to this exemplary implementation, the encoding device 10 is deployed on an entity 50 for embedding watermarks into audio signals.

[0117] The entity 50 can be, for example, a recording industry, a music producer or a service supply company providing services to the users.

[0118] The watermark can comprise data relating to signature information, copyright information, serial numbers of broadcasted audio signals, product identification in audio signal broadcasting and similar.

[0119] The audio signal can comprise music or speech as, for example, talks from movies, from TV or radio programs, from TV or radio advertisements, from TV or radio talk shows, and similar.

[0120] The decoding device 20 is deployed on a user device 60.

[0121] The user device 60 can be, for example, a PC, a smart phone, a tablet, a portable media player (e.g. an iPod®), or other similar device.

[0122] The user device 60 can be adapted to download or stream a video from the internet or to detect audio signals of a TV program broadcasted on the TV or to detect audio signals of a movie played by means of a DVD player, a VHS player, a decoder or similar.

[0123] A media provider 40 can obtain watermarked audio signals from entity 50 and supply them to user device 60, through communication network 30.

[0124] The watermarked audio signals can be, for example, supplied to the user device 60 by means of broadcasting (e.g. from a TV or radio station), streaming (e.g. from the internet) or downloading (e.g. from a PC).

[0125] The media provider 40 can be, for example, a TV station, a radio station, a PC or other similar device.

[0126] User device 60, equipped with the decoding device 20, will be configured to extract the watermark from the watermarked audio signals.

[0127] For example, the watermark can comprise information enabling the user device 60 to connect, through communication network 30, to a service provider 70 that supplies predetermined services to users. In this case, the audio signals can be, for example, audio signals of a TV talk show or movie (or similar) and the user service can involve the provision of information to users about TV images the users are watching into the TV (e.g. information about the actors, about items of clothing and/or furnishing, about the movies or talk shows, about the set and similar).

Claims

1. Method of watermarking an audio signal comprising:

- assigning the audio signal to a class, among a plurality of classes, depending on the semantic content of the audio signal, the plurality of classes being associated with a corresponding plurality of watermark profiles;
 - obtaining the watermark profile associated with the class assigned to the audio signal;
 - embedding a watermark into the audio signal by using the obtained watermark profile so as to provide a watermarked audio signal.
2. Method according to claim 1, wherein each watermark profile is associated with a corresponding class so that trade-off among watermark imperceptibility, robustness and payload is optimized for said class, depending on watermarking application.
 3. Method according to claim 1 or 2, wherein the plurality of watermark profiles relate to a single watermarking technique or to a plurality of watermarking techniques.
 4. Method according to any of claims 1 to 3, wherein each watermark profile is defined by a set of parameters and the watermark profiles differ from each other for the value taken by at least one parameter of said set of parameters and/or for at least one of the set of parameters.
 5. Method according to any of claims 1 to 4, further comprising: computing a masking threshold of the audio signal according to a psychoacoustic model.
 6. Method according to claim 5, wherein embedding the watermark into the audio signal comprises the step of shaping the energy of the watermark according to the computed masking threshold.
 7. Method according to claim 6, wherein each watermark profile is defined by a set of parameters comprising a masking threshold modulation factor F , and the energy of the watermark is shaped according to the computed masking threshold, multiplied by the masking threshold modulation factor F .
 8. Method according to any of claims 1 to 7, wherein assigning the audio signal to a class, depending on the semantic content of the audio signal, is performed based upon analysis of at least one audio signal feature related to time, frequency, energy or cepstrum domain of the audio signal.
 9. Method according to any of claims 1 to 8, further comprising a decoding process comprising: extracting the watermark from the watermarked audio signal by using the same watermark profile used for embedding the watermark into the audio signal.
10. Method according to claim 9, wherein the decoding process comprises:
 - assigning the watermarked audio signal to a class, among said plurality of classes, depending on the semantic content of the watermarked audio signal;
 - obtaining the watermark profile associated with the class assigned to the audio signal; and
 - extracting the watermark from the watermarked audio signal by using the obtained watermark profile.
 11. Method according to claim 9, wherein the decoding process comprises: trying in sequence said plurality of watermark profiles till the watermark is successfully extracted from the watermarked audio signal.
 12. Method according to claim 9, wherein embedding the watermark into the audio signal comprises: embedding a second watermark into the audio signal, comprising the class assigned to the audio signal, by using a common watermark profile; and the decoding process comprises:
 - extracting the second watermark from the watermarked audio signal by using the common watermark profile so as to retrieve the class of the audio signal,
 - obtaining the watermark profile associated with the retrieved class, and
 - extracting the watermark from the watermarked audio signal with the obtained watermark profile.
 13. System (1) of watermarking an audio signal comprising an encoding device (10) comprising:
 - a classification unit (12) configured to assign the audio signal to a class, among a plurality of classes, depending on the semantic content of the audio signal, the plurality of classes being associated with a corresponding plurality of watermark profiles;
 - a watermark profile unit (14) configured to obtain the watermark profile associated with the class assigned to the audio signal;
 - an embedding unit (16) configured to embed a watermark into the audio signal by using the watermark profile obtained by watermark profile unit so as to provide a watermarked audio signal.
 14. System (1) according to claim 13, further comprising a database (19) storing the plurality of classes associated with the corresponding plurality of watermark profiles.
 15. System (1) according to claim 13 or 14, further com-

prising a decoding device (20) configured to extract the watermark from the watermarked audio signal, by using the same watermark profile used by the embedding unit (16) for embedding the watermark into the audio signal.

5

10

15

20

25

30

35

40

45

50

55

10

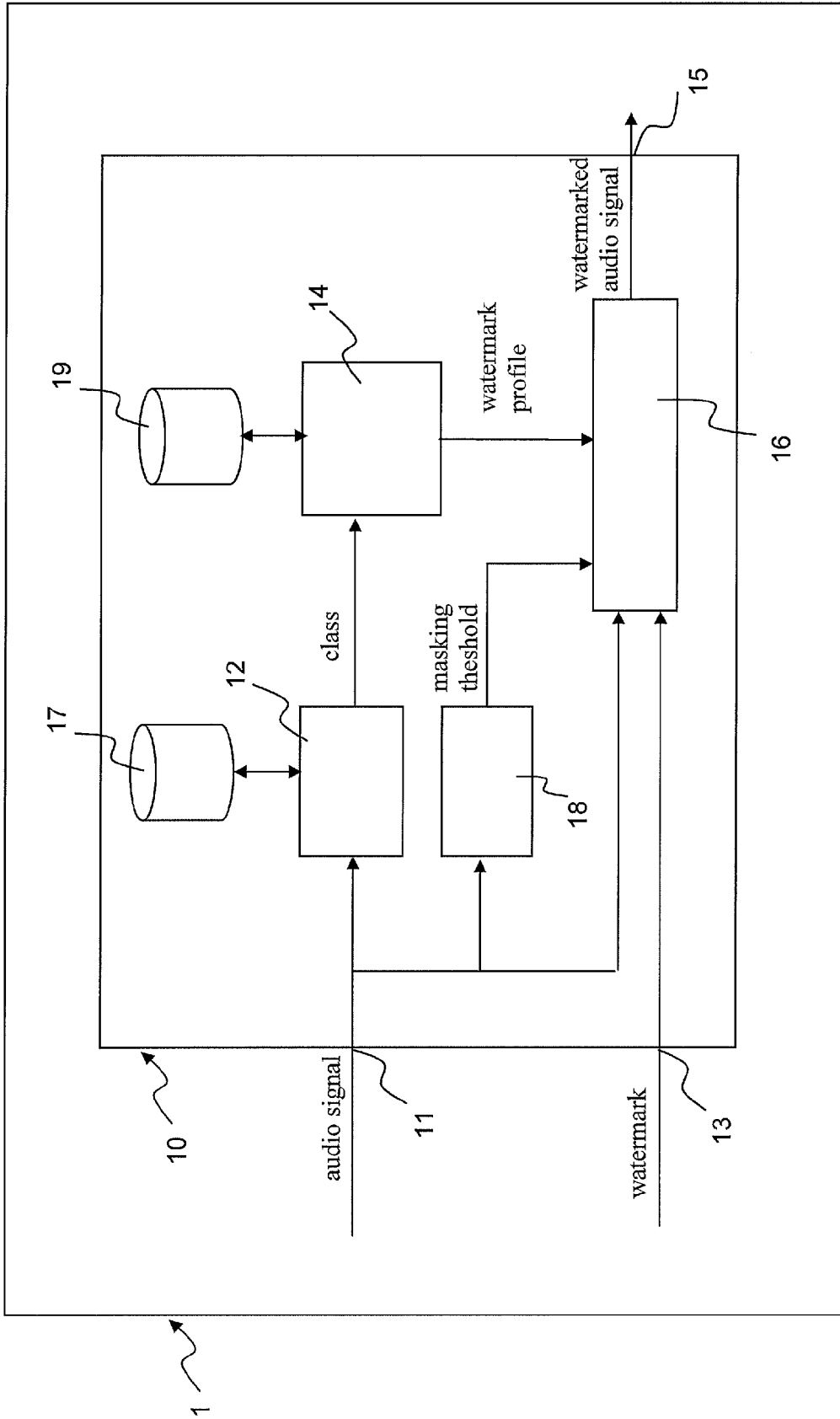


Fig. 1

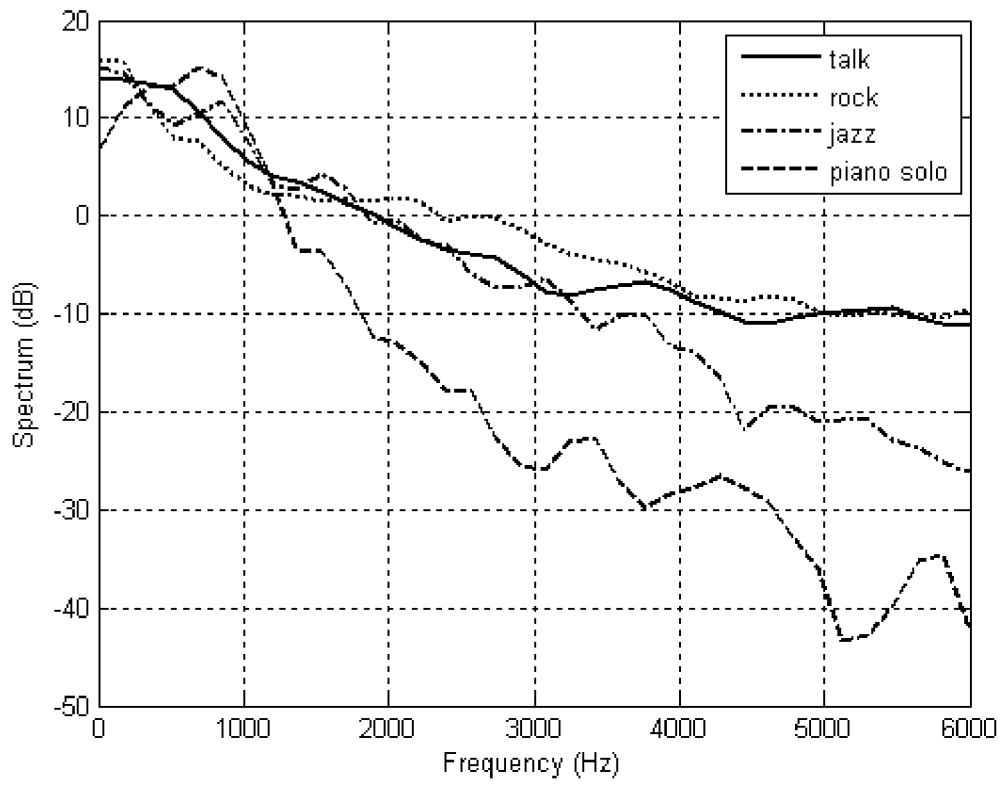


Fig. 2

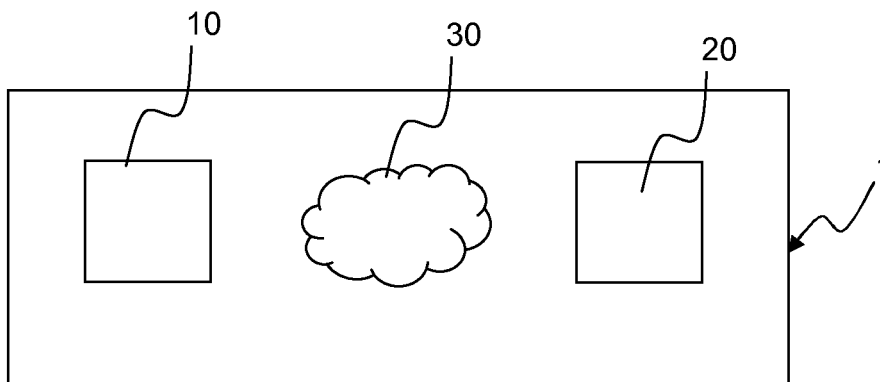


Fig. 3

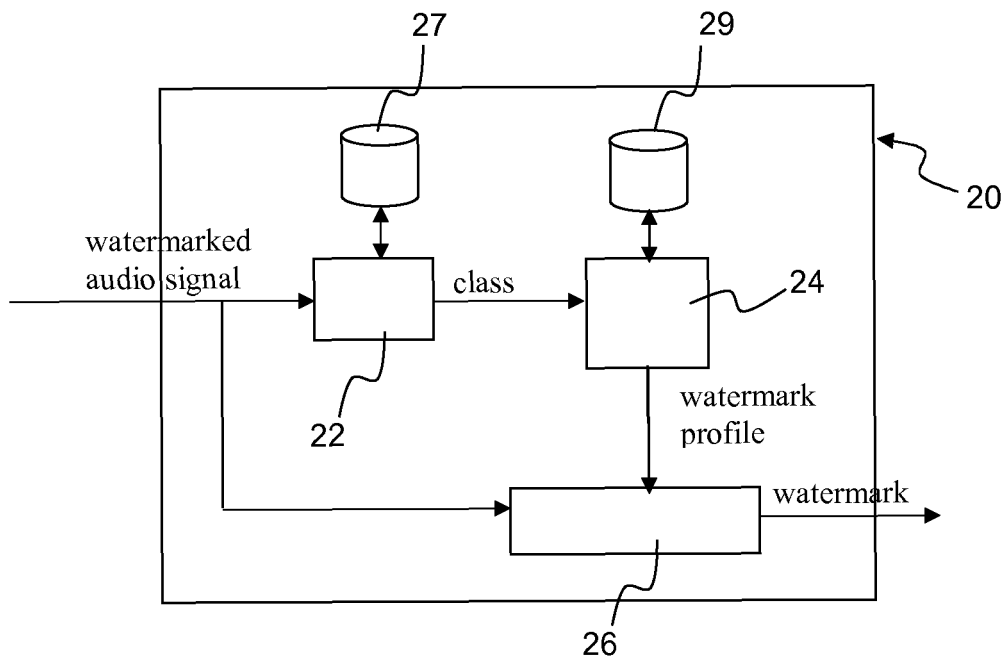


Fig. 4

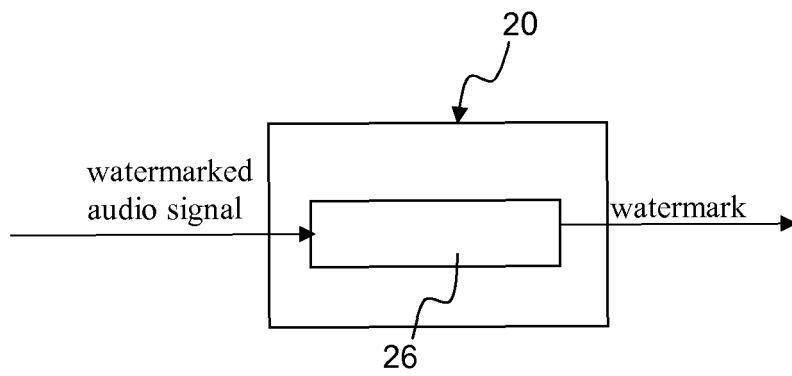


Fig. 5

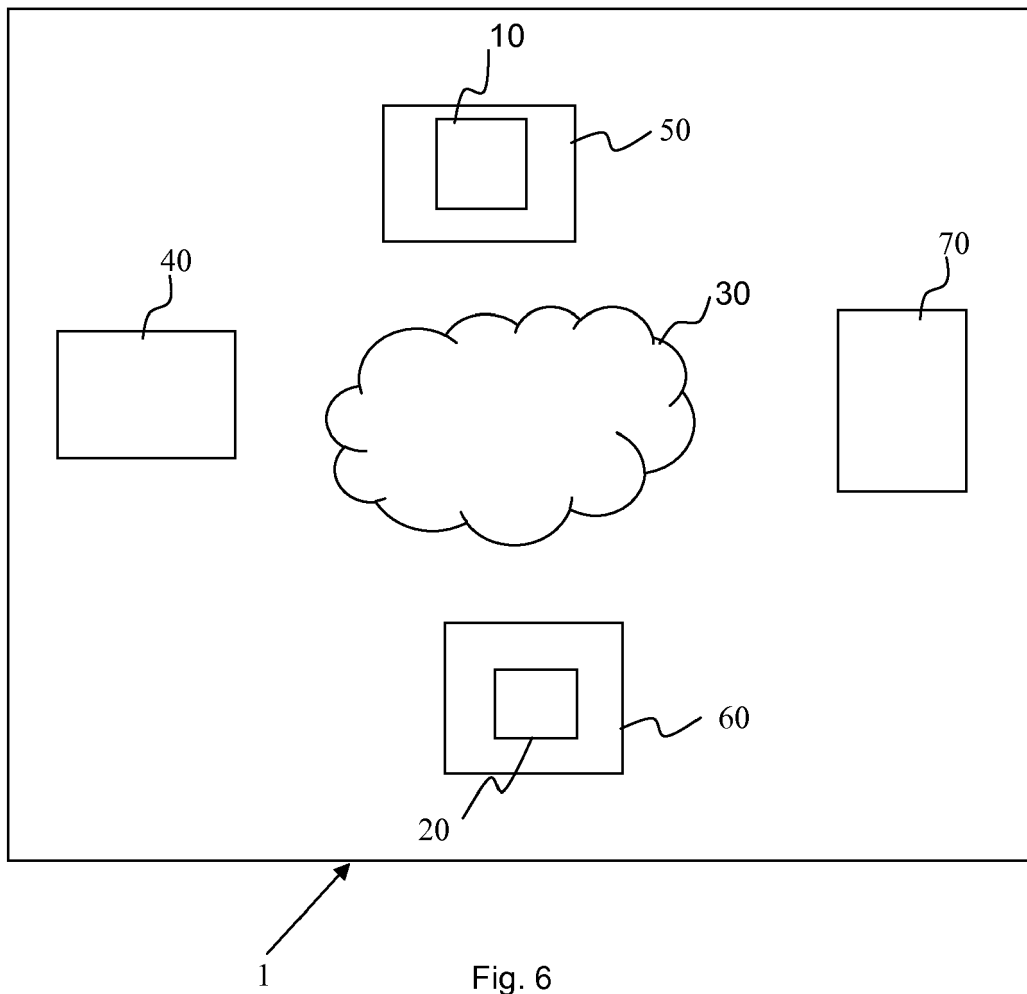


Fig. 6



EUROPEAN SEARCH REPORT

Application Number
EP 13 16 2596

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 7 127 065 B1 (DEPOVERE GEERT FLORIMOND GERAR [NL] ET AL) 24 October 2006 (2006-10-24)	1-3, 8-11, 13-15	INV. G10L19/018
Y	* column 1, line 7 - line 11 *	12	
A	* column 1, line 53 - column 2, line 17; claims 1,3 * * column 4, line 54 - column 5, line 7; claims 4,6 *	4-7	

X	US 2004/059918 A1 (XU CHANGSHENG [SG]) 25 March 2004 (2004-03-25)	1-6,8, 13,14	TECHNICAL FIELDS SEARCHED (IPC) G10L
Y	* paragraph [0008]; claim 1 *	7	
A	* paragraph [0029] - paragraph [0031]; figure 2 * * paragraph [0042] - paragraph [0043] *	9-12,15	

X	US 6 674 861 B1 (XU CHANGSHENG [SG] ET AL) 6 January 2004 (2004-01-06)	1-6,8, 13,14	G10L
Y	* column 1, line 66 - column 3, line 13 *	7	
A	* column 4, line 65 - column 6, line 10 * * column 7, line 4 - line 32 * * column 9, line 32 - column 10, line 53; figure 1 * * column 11, line 50 - column 14, line 67 *	9-12,15	

Y	US 2010/057231 A1 (SLATER CHRISTOPHER [GB] ET AL) 4 March 2010 (2010-03-04)	7	G10L
A	* paragraph [0007] - paragraph [0008] * * paragraph [0042] - paragraph [0044] *	1-6,8-15	

Y	US 6 456 726 B1 (YU HONG HEATHER [US] ET AL) 24 September 2002 (2002-09-24)	12	G10L
A	* column 1, line 6 - line 8 * * column 1, line 53 - line 62 * * column 2, line 10 - line 27; figure 1 * * column 2, line 61 - column 3, line 23 * * column 4, line 6 - line 43; figure 2 *	1-11, 13-15	

The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 13 June 2013	Examiner Virette, David
CATEGORY OF CITED DOCUMENTS		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	
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			

EPO FORM 1503_03.82 (P04C01)

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

EP 13 16 2596

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.

13-06-2013

10

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 7127065 B1	24-10-2006	CN 1338087 A	27-02-2002
		DE 60033500 T2	08-11-2007
		EP 1151413 A1	07-11-2001
		ES 2282151 T3	16-10-2007
		JP 4891508 B2	07-03-2012
		JP 2003516006 A	07-05-2003
		US 7127065 B1	24-10-2006
		WO 0139121 A1	31-05-2001
US 2004059918 A1	25-03-2004	US 2004059918 A1	25-03-2004
		WO 0249363 A1	20-06-2002
US 6674861 B1	06-01-2004	GB 2363300 A	12-12-2001
		US 6674861 B1	06-01-2004
		WO 0039955 A1	06-07-2000
US 2010057231 A1	04-03-2010	CN 101667437 A	10-03-2010
		GB 2463231 A	10-03-2010
		US 2010057231 A1	04-03-2010
US 6456726 B1	24-09-2002	CN 1399765 A	26-02-2003
		DE 60012992 D1	16-09-2004
		DE 60012992 T2	08-09-2005
		EP 1256086 A1	13-11-2002
		JP 2004500677 A	08-01-2004
		US 6456726 B1	24-09-2002
		WO 0131565 A1	03-05-2001

20

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

- EP 1594122 A [0005]