



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
23.11.2011 Bulletin 2011/47

(51) Int Cl.:
G10L 21/04 (2006.01)

(21) Application number: **11155349.1**

(22) Date of filing: **22.02.2011**

(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

(30) Priority: **19.05.2010 US 346124 P**

(71) Applicant: **Fraunhofer-Gesellschaft zur Förderung der Angewandten Forschung e.V.**
80686 München (DE)

(72) Inventors:
• **Nagel, Frederik**
90425 Nürnberg (DE)
• **Geyersberger, Stefan**
97076 Würzburg (DE)
• **Disch, Sascha**
90766 Fürth (DE)
• **Neuendorf, Max**
90402 Nürnberg (DE)

(74) Representative: **Zinkler, Franz**
Schoppe, Zimmermann, Stöckeler & Zinkler
Patentanwälte
Postfach 246
82043 Pullach bei München (DE)

(54) **Apparatus and method for extending or compressing time sections of an audio signal**

(57) An audio signal processor (100) comprising an analysis means (104), a manipulation factor unit (106), and a time-stretching and compression device (108). The analysis means (104) is implemented to determine a first measure of information content (M_1) of a first time section of an audio signal and a second measure of information content (M_2) of a second time section. The manipulation factor unit (106) is implemented to determine a time manipulation factor (ΔD_1) for the first time section in dependence on the first measure of information content (M_1) and the second measure of information content (M_2). The time-stretching and compression device (108) is implemented to time-stretch or compress the first time section according to the manipulation factor (ΔD_1) and to treat the second time section differently from the first time section. A corresponding method for adjusting time information content variations of an audio signal (s) is also disclosed. The determination of the first measure of information content and of the second measure of information content may be based on externally provided control information such as meta-data provided along with the audio signal.

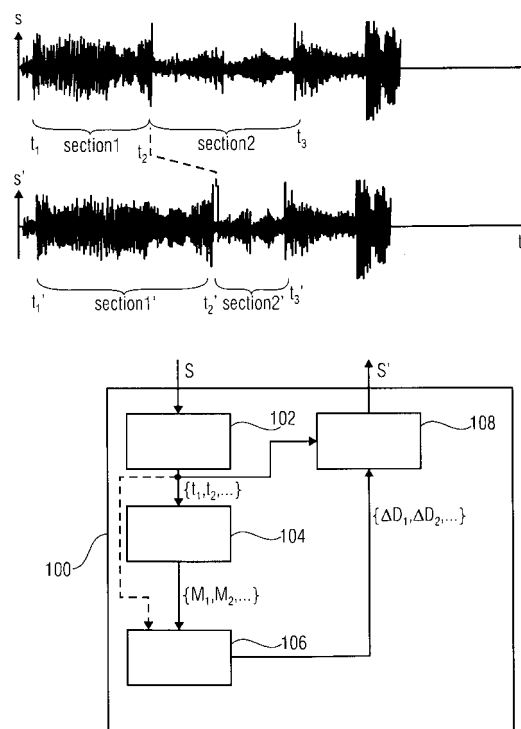


FIG 1

Description

[0001] Embodiments according to the invention relate to audio processing and particularly to an apparatus and a method for extending or compressing an audio signal in a time section-wise manner.

[0002] Recorded audio signals may be played back at a speed different from the original speed at which the (original) audio signal was recorded. This may be useful to slow down or accelerate the audio signal so that a listener may receive the information conveyed by the audio signal at a rate that is convenient to the listener. The listener may, for example, choose a relatively fast playback speed when seeking a certain segment within the audio signal by paying attention to a particular keyword used within the segment sought. Typically, the listener will not be able to mentally process the entire information conveyed by the audio signal if the audio signal is played back at a fast speed. Nevertheless, the keyword typically can be recognized by the listener, even at relatively high playback speeds, if the listener concentrates on detecting the keyword. Another option is to choose a slower speed for playback which is useful when the listener wants to extract relevant information from the audio signal. For example, the audio signal may have been recorded during a court hearing and a transcript of the audio signal needs to be prepared. Yet another example can be found in the field of aviation when the audio signal has been extracted from a flight data recorder, wherein specialists are commissioned with the identification of various utterances and sounds that can be heard when playing the audio signal. Varying the playback speed of the audio signal may facilitate the identification of the various recorded sounds.

[0003] Currently known methods for temporarily extending and compressing an audio signal receive control information from a user, e.g. the listener. For example, a time-stretching method has been introduced under the name "Phavorit" and described in an article by T. Karrer, E. Lee, and J. Borchers, "PhaVoRiT: A Phase Vocoder for Real Time Interactive Time-Stretching" in Proc. ICMC, Nov. 2006, pp.708-715". This time-stretching method is based on a phase Vocoder for which the user may select the degree of time-stretching at runtime.

[0004] This and other currently known methods process an entire audio signal in an across-the-board manner or they have to be explicitly controlled by the user.

[0005] The alternation of the playback speed of an audio signal typically causes a change of the pitch of the audio signal. If this is not desired, many different time stretching methods may be used, such as synchronous overlap and add (SOLA), pitch synchronous overlap and add (PSOLA), waveform similarity overlap and add (WSOLA), Pointer Interval Controlled Overlap and Add (PICOLA), Time Domain Harmonic Scaling (TDHS), Minimum Perceived Loss Time Compression/Expansion (MPEX), or the phase vocoder. Each of these techniques has some advantages for certain signals. The following,

however, concentrates on the phase vocoder.

[0006] In an article "The Phase Vocoder: A Tutorial" by Mark Dolson, Computer Audio Research Laboratory, Center for Music Experiment, University of California, San Diego, the operation of the phase vocoder is explained. The phase vocoder is a signal processing technique (typically digital) that can be used to perform very high fidelity time-scaling, pitch transposition and other modifications of recorded sound.

[0007] In the German Patent Application publication DE 10 2008 015 702 A1, a device and a method for bandwidth expansion of an audio signal is described. The device uses a phase vocoder in filterbank implementation or transformation implementation for temporarily spreading the audio signal by a predetermined, constant factor.

[0008] It would be desirable to provide a device and a method for automatically and selectively stretch and/or compress individual sections of audio signals, in particular, speech signals. This operation may be carried out with either SOLA, WSOLA, PSOLA, PICOLA, TDHS, MPEX, phase vocoder or other time or pitch scaling techniques.

[0009] This desire and/or other desires are addressed by an audio signal processor according to claim 1, a method according to claim 14, or a computer program according to claim 15.

[0010] An embodiment of the invention provides an audio signal processor which comprises an analysis means, a manipulation factor unit, and a time-stretching and compression device. The analysis means is implemented to determine a first measure of information content of a first time section of an audio signal and a second measure of information content of a second time section. The manipulation factor unit is implemented to determine a time manipulation factor for the first time section in dependence on the first measure of information content and the second measure of information content. The time-stretching and compression device is implemented to time-stretch or compress the first time section according to the manipulation factor and to treat the second time section differently from the first time section.

[0011] By applying different manipulation factors regarding the time stretching and compression to different time sections, time sections having a higher measure of information content (e.g. a higher information density), can be time-stretched or temporarily extended. In the alternative, time sections having a relatively low measure of information content may be temporarily compressed or even deleted from the signal. The audio signal processor also facilitates a combination of both options. With the proposed audio signal processor it is possible to distribute information content more evenly over the duration of the audio signal.

[0012] In the related field of perceptual speech and audio coding, current methods of speech and audio coding may not code signal components that are perceived as noise-like, but synthesize an equally noise-like perceived signal at the receiver side, possibly using a few

parametric values transmitted from the sender to the receiver. This receiver side substitution is typically limited to noise. This technique is called Perceptual Noise Substitution (PNS). The substituted signal components typically are not dispensable, but contain e.g. sibilant sounds etc. with a high semantic content.

[0013] Another technique used in mobile phones is the insertion of comfort noise. The purpose of this technique is to reduce the amount of data that needs to be transmitted or stored, especially in the case of noise. In contrast, the proposed audio signal processor makes it possible to use the noise-filled time section as a freed-up resource that is usable for other information. This inherently helps to maintain the quality and/or intelligibility of important signals portions, while less effort is spent for the coding of noise like segments of an audio signal.

[0014] This functionality of the audio signal processor is not limited to noise or noise-like signal components, but also to other signal components having a low measure of information content. The choice of what kind of signals qualify as having a relatively high measure of information content is a question of implementation, configuration, and user preferences. This question and its solution are typically addressed during an implementation process of the analysis means. Another difference to the above mentioned current methods of speech and audio coding is that with the proposed method and apparatus a noise-filled time section is not filled with a synthesized noise that is modeled to imitate the original noise. Instead, the noise-filled time section and other time sections having a low measure of information content may be filled with payload information. With PNS-based methods, the noise is re-synthesized at the decoder side. However, irrelevant speech segments and pauses are not considered separately. Furthermore, these methods regarding audio coding do not modify durations of time sections within the audio signal or the duration of the complete audio signal, because this would be contradictory to the goal of the audio coding methods, which is to achieve a high degree of similarity between the original signal at the coder side and the decoded signal at the decoder side.

[0015] The determination of the first measure of information content and of the second measure of information content may be based on externally provided control information, that is, the analysis means may be configured to identify and/or extract the control information. For example, the analysis of the externally provided control information may be performed in real time on the basis of information additionally supplied along with the audio signal. The information content data could be sent along as meta-information or meta-data, which indicates the information content of one or several time sections.

[0016] Current methods for time-stretching and compressing of audio signals act on an audio signal in a global, across-the-board-manner which results in pauses and superfluous sections being time-stretched or compressed at the same time and to the same degree as

other time sections. Audio coding methods that exploit the irrelevance of information, do so either exclusively with respect to masking or with respect to noise.

[0017] The audio signal processor according to the teachings disclosed herein extracts parameters regarding a degree of (time section-wise) stretching and compressing from the signal itself, which does not appear to be implemented in currently known methods.

[0018] The audio signal processor may be used to automatically detect or estimate speech pauses and to use these time intervals for a selective speech stretching. The words may be played back at a slower speed. In turn, the pauses are shortened. Primarily, pauses designate a being silent of the speaker. However, the term may be extended to so-called "filled" pauses. Filled pauses refer to makeshift words such as "er", "um", "well", etc., or the repetition of words and word parts, for example, "I mean mean mean...". Stuttered syllables fall into this category as well. All of these pauses have in common that they do not contain information in the sense of exchanging facts and are thus substantially negligible as irrelevant. In the literature, these pauses are sometimes referred to as "filled pauses".

[0019] The time stretching of selected time sections may improve the comprehensibility of the audio signal and enable non-native speakers, aurally challenged persons, senior citizens etc. to follow spoken texts more easily. In addition, such detection could be used in audio or speech coding, as filled pauses may be coded at an inferior quality or not at all.

[0020] In some embodiments of the teachings disclosed herein, the audio signal processor may further comprise a comparator implemented to compare the first measure of information of the content of the first time section to a threshold and to classify the first time section, in dependence on a respective result of the comparison, as a section having a higher measure of information content or as a section having a lower measure of information content. A section bounding means may be provided that is implemented to shift boundaries between the sections having the higher measure of information content and the sections having the lower measure of information content into or towards the sections having lower information content. The time-stretching and compression device may further be implemented to time-stretch or compress the sections having a higher measure of information content by a factor corresponding to the shift of the boundaries of the first time section.

[0021] Another embodiment of the teachings disclosed herein provides a method for adjusting time information content variations of an audio signal which comprises: determining a first measure of information content of a first time section of the audio signal and a second measure of information content of a second time section of the audio signal; determining a time manipulation factor for the first time section in dependence on the first measure of information content and the second measure of information content; and processing the audio signal such

that the first time section is time-stretched or compressed according to the time manipulation factor and that the second time section is processed differently from the first time section.

[0022] The dependent claims relate to further enhancements and/or details of the audio signal processor, the method for adjusting time information content variations of the audio signal, and/or the computer program.

Brief Description of the Drawings

[0023] The accompanying drawings are included to provide a further understanding of embodiments and are incorporated in and constitute a part of this specification. The drawings illustrate embodiments and, together with the description, serve to explain the principles of the embodiments. Other embodiments and many of the intended advantages of embodiments will be readily appreciated, as they become better understood with reference to the following detailed description. Like reference numerals designate corresponding similar parts.

- Fig. 1 is a schematic block diagram of an audio signal processor according to the teachings of this document;
- Fig. 2 is a schematic block diagram of another embodiment of an audio signal processor according to the teachings of this document;
- Fig. 3 is a diagram illustrating measures of information content for a plurality of time sections of the audio signal over time;
- Fig. 4 shows another diagram of measure of information content over time illustrating a hysteresis concept;
- Fig. 5 is a schematic block diagram of another embodiment of an audio signal processor according to the teachings disclosed herein;
- Fig. 6 is a schematic block diagram of another embodiment of an audio signal processor according to the teachings disclosed herein;
- Fig. 7 is a schematic flowchart of an embodiment of a method for adjusting time information content variations of an audio signal according to the teachings disclosed herein;
- Fig. 8 shows a flowchart of another embodiment of a method for adjusting

time information content variations,

Figs. 9a to 9e show diagrams of an energy quantity of the audio signal over time to illustrate various actions of an embodiment of the method for adjusting time information content variations; and

Figs. 10a to 10e show diagrams of a time manipulation factor over time for different implementations or configurations of the audio signal processor and/or the method for adjusting time information content variations.

Detailed Description of the Embodiments

[0024] Fig. 1 shows a schematic block diagram of an audio signal processor 100 according to an embodiment of the teachings of this document. The audio signal processor 100 receives an audio signal s as an input. The audio signal s is illustrated at the top of Fig. 1 in a diagram of a signal amplitude over time. The audio signal contains relatively large amplitude values in a first time section (section 1), which extends between two time instants t_1 and t_2 . Without loss of generality and as an illustrative example, it is assumed that the first section contains relevant information in the form of spoken words and therefore has a high measure of information content. A second time section (section 2) extends between the time instants t_2 and t_3 . On average, the amplitude values are lower in the second section than in the first section. For the purpose of illustration, it is assumed that this indicates a low measure of information content for the second time section.

[0025] Within the audio signal processor 100, the audio signal s is input to a section identifier 102. The section identifier 102 may perform a coarse analysis of the audio signal s to determine changes in characteristic properties of the audio signal s from one time instant to another. Large changes may be an indicator of boundaries between two time sections of the audio signal s . A simpler implementation of the section identifier 102 splits the audio signal s into time sections of equal length (e.g. between 1/10 seconds to several seconds). Other implementations of the section identifier 102 are also possible. The section identifier 102 produces a set of time instant values $\{t_1, t_2, \dots\}$ to be used by other components of the audio signal, processor 100.

[0026] The audio signal s may be provided to the audio signal processor as a digital, pulse code modulated (PCM) signal. Other forms for the audio signal s are also possible, and even an analog representation. In the case of s being an analog signal, it would either be analog-to-digital converted for subsequent digital signal processing or it would be analyzed and processed as an analog signal.

[0027] The set of time instants $\{t_1, t_2, \dots\}$ is transmitted

to an analysis means 104 in the form of, for example, a vector or a list. The analysis means processes the audio signal s in a time section-wise manner to determine a plurality of measures of information content for the plurality of time sections. Thus, the analysis means 104 determines a high measure of information content for the first section shown in the time diagram for the audio signal s , and a lower measure of information content for the second section. The measures of information content are indicated by reference signs $M_1, M_2, \dots$.

[0028] In order to determine and quantify the measure of information content for a given time section, the analysis means 104 may analyze the audio signal within the time sections in a variety of ways. A relatively simple implementation is based on evaluating the strength of the audio signal within the given time sections. To this end, an average amplitude or power of the audio signal within the given time sections may be determined. The determination of a maximal value within the given time section is another option. An amplitude-based or power-based analysis of the audio signal is suitable to distinguish between silent parts and non-silent parts of the audio signal. A more complex approach is to perform a spectral analysis of the time section to find out how the audio signal is distributed in the frequency-domain. A relatively uniform distribution of the audio signal's power spectrum over the frequency range occupied by the audio signal could indicate that the audio signal mostly consists of noise in the evaluated time section. Yet another option for the implementation of the analysis means 104 is given by a pattern detection. The audio signal in the time section is compared with a plurality of sound samples and the most similar sound samples are retained. Each sound sample may have an information associated with it indicating the nature of the sound sample, e.g. high measure of information content or low measure of information content. A more elaborate approach could even distinguish between, e.g. a man's voice, a woman's voice, a child's voice, noise, traffic noise, etc. Based on the result of the comparison, the analysis means may determine whether the time section in question has a high measure of information content or a low measure of information content. As an option, the analysis means 104 may locally measure the speech velocity (e.g. syllables rate), for determining whether a time section of the audio signal s essentially comprises spoken language, and if applicable, for determining a speech velocity within the time section. The information about the speech velocity may be used for controlling the time-stretching and/or compression of individual time sections within the audio signal s . Another option for the analysis means 104 is to receive externally provided control data, for example as meta-information provided along with the audio signal s .

[0029] The set of measures of information content $\{M_1, M_2, \dots\}$ is transmitted to a manipulation factor unit 106. The manipulation factor unit 106 determines a plurality of manipulation factors $\{\Delta D_1, \Delta D_2, \dots\}$ (the letter D stands for "duration"). For example, the manipulation factor unit

106 may assign a manipulation factor ΔD_i resulting in a time-stretch to be performed on a corresponding time section i if the corresponding measure of information content M_i is high. In contrast, time sections with a low measure of information content are assigned a manipulation factor that results in a compression of the corresponding time section. The manipulation factor unit 106 may optionally receive the set of time instants $\{t_1, t_2, \dots\}$, too. Based on the information about the time instants of the boundaries between the various time sections, the manipulation factor unit 106 can evaluate how much margin is available in time sections with low measure of information content that can be used for time-stretching adjacent time sections having a higher measure of information content. This may be useful if it is intended that the time-stretching and compression will not modify the temporal positions of the corresponding time section within the entire audio signal and/or a total duration of the entire audio signal. For example, consider the case where the audio signal is an audio track of a movie picture. Assuming that one time section corresponds more or less to one line of an actor, it is important that the time section of the audio signal is played back substantially at the same time as the image of movie picture shows the actor pronouncing the line. Although a perfect synchronization of the played back audio signal is typically no longer feasible due to the time-stretching or compression of the time section, at least the beginning of the actor's line can be synchronized to the image of the movie picture so that the viewer knows what the actor said during a specific scene. Thus, the audio signal processor and the manipulation factor unit in particular may be implemented to preserve the temporal position of a given time section in the audio signal with respect to a beginning, an end, or a center of the given time section.

[0030] The set of manipulation factors $\{\Delta D_1, \Delta D_2, \dots\}$ is sent from the manipulation factor unit 106 to a time-stretching and compression device 108 in the form of e.g. a vector, a list, or a handover in one or more registers of the audio signal processor 100. The time-stretching and compression device 108 also receives the set of time instants $\{t_1, t_2, \dots\}$ from the section identifier 102, so that the time-stretching and compression device 108 can perform time-stretching and/or compression operations at the intervals indicated by the time instants provided from the section identifier 102. Time-stretching and compression may be done by resampling the audio signal at a higher or lower sampling rate. The resampled audio signal is then decimated or interpolated in order to obtain the original sampling rate again. Resampling and decimating or interpolating the audio signal typically causes a modification of the pitch of the audio signal in the affected time section. The modification of the pitch may be used as an indicator to the listener by how much a particular time section has been time-stretched or compressed. If the modification of the pitch is not desired, it may be prevented, e.g., by using a phase vocoder. The phase vocoder provides a high quality solution for time-

scale modification of signals. Pitch-scale modifications are usually implemented as a combination of time-scaling and sampling rate conversion. For a detailed description of phase-vocoders, reference is made to the following citations:

- "The Phase-Vocoder: A Tutorial", Mark Dolson, Computer Music Journal, Vol. 10, No. 4, pages 14-27, 1986;
- "New Phase-Vocoder Techniques for Pitch-Shifting, Harmonizing and Other Exotic Effects", Jean Laroché and Mark Dolson, Proceedings 1999 IEEE, Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, Oct. 17-20, 1999, pages 91-94;
- "New Approach to Transient Processing in the Phase Vocoder", A. Röbel, Proceedings of the International Conference on Digital Audio Effects of DAFx-03, London, UK, September 8-11, 2003, pages DAFx-1 to DAFx6;
- "Phase-locked Vocoder", Meller Puckette, Proceedings 1995 IEEE, ASSP, Conference on Applications of Signal Processing to Audio and Acoustic Noise;
- US Patent Nr. 6,549,884.

[0031] Other available methods and techniques to time-stretch and/or compress time sections of the audio signal are provided by the PSOLA, WSOLA, SOLA, PICOLA, TDHS, and MPEX method.

[0032] The output of the time-stretching and compression device 108 and typically also of the audio signal processor 100 is a modified audio signal s' , as illustrated in the second time diagram of Fig. 1. It can be seen that the first section in the modified audio signals (section 1') has been time-stretched at the cost of the second time section (section 2'). This leads to a shift of the boundary between the first section and the second section t_2 to a new value t_2' . The time instants t_1' and t_3' are substantially unchanged and therefore substantially equal to t_1 and t_3 , respectively. Note however that, in departure from the illustration in Fig. 1, the time section to the right of section 2 might also have been subjected to a time-stretching operation. In that case the time instant t_3 would have been shifted to the left, so that the time interval for section 2 would be even stronger compressed.

[0033] Fig. 2 shows another embodiment of the audio signal processor according to the teachings of this document. The section identifier 102, the analysis means 104, and the time-stretching and compression device 108 are substantially identical to the ones described in Fig. 1. The analysis means 104 provides the set of measurements of information content $\{M_1, M_2, \dots\}$ to a comparator 204 which compares the measure of information content to a threshold M_{thr} and classifies each one of the plurality

of time sections as having a high(er) measure of information content or a low(er) measure of information content. Thus, two classes are formed which reflects the fact that a time section can be either time-stretched or compressed. A third possibility would be to leave some of the time sections unaltered, giving rise to a potential third category and a second threshold for the measurement of the information content.

[0034] The threshold(s) may either be predetermined and fixed or variable in order to be adapted to the properties of a given audio signal. For example, one strategy may be to determine the threshold M_{thr} in a manner that the number of sections with high measure of information content is approximately equal to the number of time sections with low measure of information content. Thus, a relatively high number of boundaries between sections of different information content measure are obtained which increases the degrees of freedom for the manipulation factor unit 106 and/or the time-stretching and compression device 108. To this end, all of the measures of information content would be determined in a first step, then sorted according to their respective measures of information content, and finally the threshold would be set to the mean measure of information content.

[0035] The comparator 204 produces a set of classification values $\{C_1, C_2, \dots\}$ which is provided to a section bounding means 206. The section bounding means 206 is implemented to shift boundaries between the section having a higher measure of information content and the sections having a lower measure of information content for the benefit of the former time sections and at the cost of the latter time sections. For example, the boundaries are shifted into the time sections having the lower measure of information content. These section bounding means 206 further receive the set of time instants $\{t_1, t_2, \dots\}$ from the section identifier 102. The set of time instants is also supplied to the manipulation factor unit 106, as well as a set of shifted time instants $\{t_1', t_2', \dots\}$ by determining the difference between the original time instants and the shifted time instants. The manipulation factor unit 106 can determine the time-stretch or compression factors for the various time intervals. The determined manipulation factors are again transmitted to the time-stretching and compression device 108.

[0036] Fig. 3 shows a diagram of the measurement of information content determined for a plurality of time sections. The measure of information content M is piecewise constant for the duration of at least one time section in this embodiment. The measure of information content is compared to a threshold M_{thr} . Based on a result of the comparison, the time sections are classified as sections having a higher measure of information content or as sections having a lower measure of information content. Two or more adjacent sections having the same classification may be combined to a contiguous region of time sections. For the purposes of time-stretching and compressions, the contiguous regions may be regarded as a unit, e.g. the same manipulation factor applies for all

time sections within the contiguous region. The boundaries between adjacent contiguous regions are determined and shifted by an amount depending on the time manipulation factors valid for the two adjacent contiguous regions typically into the one of the contiguous regions that has the lower measure of information content. Time-stretching or compressing the first time section comprises time-stretching or compressing the time sections making up a contiguous region having a higher measure of information content into at least one adjacent contiguous region having a lower measure of information content, in correspondence to the shifted boundary and the amount of shifting.

[0037] Fig. 4 shows a graph similar to the one shown in Fig. 3 of a measure for information content M as determined for a plurality of time sections. Depending on the length of the time sections, in particular, when the length of the time sections and/or the threshold M_{thr} is predetermined and fixed, a certain audio signal may cause a lot of transitions from time sections with low measure of information content $M_i = LO$ to time sections with a high measure of information content $M_i = HI$. This may, for example, happen when the audio signal has been recorded at a low recording level or the speaker has spoken in a relatively soft voice. In order to avoid too rapid changes between high and low information content measure sections, the comparator 204 may be arranged to determine the classification result using a hysteresis. As can be seen in Fig. 4, the comparator 204 uses two threshold values, M_{hi} and M_{lo} . A boundary between a preceding time section with a low measurement of information content and a subsequent time section with a high measure of information content occurs if the higher threshold M_{hi} is exceeded in the upward direction. On the other hand, a transition from a time section with high information content measure to one with low information content measure occurs when the lower threshold M_{lo} is exceeded in the downward direction. Thus, the contiguous regions resulting from combining several adjacent time sections are larger than without a hysteresis. It can thus be avoided that the audio signal is split up into too many contiguous regions, leading to a high number of manipulation factors. It may be confusing to the listener if the manipulation factor changes too frequently.

[0038] The choice of, and interaction between, the values for the thresholds M_{thr} , M_{hi} , M_{lo} and the length of the elementary time sections may be subject to a preprocessing step in which the audio signal is evaluated with respect to e.g. an average level of information content.

[0039] Fig. 5 shows a schematic block diagram of another embodiment for an audio signal processor 100 according to the teachings of this document. The audio signal processor 100 now further comprises a limiting device 508 for the time-stretching or compression. The limiting device 508 is implemented to determine a current threshold for the time stretching or compression of the section having higher information content and to limit the time stretching and compression to the current threshold. Fig.

5 shows an embodiment in which the limiting device 508 implements an upper threshold ΔD_{max} , and a lower threshold ΔD_{min} . In the interval $[\Delta D_{min}, \Delta D_{max}]$, the limiting device 508 is substantially a unity function, i.e. an output of the limiting device 508 is substantially equal to an input thereof. Outside this interval, the output value is limited to the respective lower or upper value. The output of the limiting device 508 is a set of limited manipulation factors $\{\Delta D_1', \Delta D_2', \dots\}$. The limiting device 508 and a corresponding limiting action of a method for adjusting time information content variations of the audio signal avoids time-stretching or compressing the time sections of the audio signals s with excessive manipulation factors which would result in a speech being played back too slowly or too fast, for example.

[0040] Fig. 6 shows a schematic block diagram of another embodiment of an audio signal processor 100 according to the teachings disclosed herein. The audio signal s is also supplied to a speech velocity measuring device for determining whether a time section of the audio signal s essentially comprises spoken language, and if applicable, for determining a speech velocity within the time section. A set of section related speech velocity measures $\{v_1, v_2, \dots\}$ is output by the speech velocity measuring device 602 and forwarded to a threshold setting device 608. The threshold setting device 608 is connected to the speech velocity measuring device 602 and intended for determining, based on the determined speech velocity, at least one threshold for the manipulation factor valid for the time section in question. The threshold setting device 608 is further connected at an output of the threshold setting device 608 to the limiting device 508. The limiting device 508 receives a current threshold value or several current threshold values ΔD_{max} and ΔD_{min} from the threshold setting device 608.

[0041] The embodiment of the audio signal processor 100 shown in Fig. 6 can be used for controlling the time-stretching and/or compression of individual time sections within the audio signal s . In particular, the degree of time-stretching and/or compressing can be determined as a function of an instantaneous speech velocity. By controlling the audio signal processing via an estimate of the instantaneous speech velocity, a balanced, substantially uniform speech velocity may be obtained over the entire speech signal due to such processing. This may be particularly helpful with intermittently performed speech or an irregular speech velocity. The speech comprehensibility of such speech presentations thus may be improved.

[0042] The set of estimated speech velocities $\{v_1, v_2, \dots\}$ may also be supplied directly to the manipulation factor unit 106 instead of to the threshold setting device 608, or in addition thereto. It is also possible to use the speech velocity estimate as a measure of information content in the various time sections, or as a precursor thereof. In this case, the speech velocity measuring device 602 may be a part of the analysis means 104.

[0043] In the context of the method for adjusting time

information content variations of the audio signal s , the following actions may be performed in connection with a speech velocity measurement:

- determining whether the audio signal s essentially comprises spoken text within a given time section;
- determining a speech velocity of the spoken text during the given time section when the audio signal s essentially comprises spoken text within the given time section; and
- determining, in dependence on the speech velocity, at least one threshold for the manipulation factor for the given time section.

[0044] One method that may be used to determine or estimate the speech velocity is to detect phonemes in the audio signal s and to count the number of phonemes per time unit. Per definition, a phoneme is the smallest segmental unit of sound employed to form meaningful contrasts between utterances in a language or dialect.

[0045] Fig. 7 shows a schematic flowchart of an embodiment of the method for adjusting time information content variations in the audio signal s . The method illustrated by the flowchart comprises some optional actions that are not part of a base embodiment of the method. After the start of the method, first and second measures of information content are determined, the first measure of information content corresponding to a first time section of the audio signal s and the second measure of information content corresponding to a second time section of the audio signal s (reference number 702). As shown in the box with reference number 704, at least the first measure of information content may be compared with a threshold M_{thr} . Typically, the measures of information content M_i of all time sections are compared with the threshold. The comparison of the information content measures with the threshold M_{thr} is a preparatory action to a classification of the one or more time sections as a section with a high measure of information content or as a section with a low measure of information content (reference sign 706). Alternative embodiments may use three or more classes instead of only the two classes for high and low information content. The grouping of time sections having approximately equal measures of information content into a countable number of classes makes it possible to combine adjacent time sections having equal classification results, that is being in the same class, to form larger contiguous regions within the audio signal class, in which the measure of information content is approximately constant. Such a contiguous region may correspond to e.g. a complete sentence spoken by a speaker without any significant pauses. The combination of the adjacent time sections is represented by the box 708 in the flowchart of Fig. 7.

[0046] In this embodiment of the method, the measures of information content are determined for relatively

short time sections (on the order of a fraction of a second to a few seconds, e.g. 0.5 seconds, 1 second, 2 seconds, or 5 seconds). Thus, a relatively fine granularity can be achieved which facilitates a relatively precise detection of time instants in the audio signal s , at which the measure of information content varies significantly, such as at the end of a spoken passage followed by a pause or silence. On the other hand, the contiguous regions are typically larger than a single time section and thus allow the time-stretching or compressing of longer passages.

[0047] At 710 boundaries between the adjacent contiguous regions are determined, and then, at 712 a security zone is inserted in the contiguous regions having a low measure of information content. The security zone is typically inserted adjacent to the boundary to the time section with a high measure of information content. This will be explained in more detail below in the context of Fig. 9c. In short, the insertion of the security zone is done to prevent the beginnings and ends of spoken passages to be treated as having a low measure of information content only, which may occur due to edge effects or certain phenomena of spoken language occurring at the beginning or the end. The security zone is then attached to the adjacent region having a high measure of information content. Thus, the security zone will be treated as a part of the high information content measure or region, i.e. it undergoes the same time-stretching and/or compression (cf. reference sign 714).

[0048] A time manipulation factor is determined for the first time section in dependence on the first measure of information content, and the second measure of information content at 716. The determination of the manipulation factor ΔD_i may evaluate how much resource in the form of time intervals having a low measure of information content, is available around a time section having a high measure of information content, so that the high information content section may be time-stretched into the low information content section. When time sections containing substantial pauses or silence are compressed for the benefit of time sections containing spoken language, the determination of the time manipulation factor may keep a shorter pause or silence which may help, for example, a listener to mentally segment two subsequent sentences from each other.

[0049] A currently valid threshold ΔD_{max} , ΔD_{min} for the time manipulation factors ΔD_i is determined at an action having the reference sign 718 in Fig. 7. Then, at 720, the time manipulation factor ΔD_i for a given time section is limited according to the current threshold ΔD_{max} , ΔD_{min} .

[0050] The audio signal s is then processed by time-stretching or compressing the first time section(s) as indicated by action 722 in Fig. 7. It is to be noted that the method may be repeated or that only selected actions of the method may be repeated.

[0051] Fig. 8 shows another schematic flow diagram of another embodiment of the method for adjusting time information content variations. The speech signal s is supplied to a pause detection 802 and to an optional filled

pause detection 804. Filled pauses contain less important information, such as make-shift words (mh, oh, well, etc), or repeated words, to name a few. At an action 806, the pauses are at least partially removed. The pause removal may comprise a determination of modified time instants in the audio signal s to which non-pause time sections of the audio signal s may be extended. A result of the pause removal action 806 is supplied to a function block 818 charged with the creation of a time-stretching function. Both, the pause removal 806 and the time-stretching function 818 are controlled by control parameters 808, such as thresholds. The time-stretching function 818 is then applied to the audio signal s at 822, which yields the modified audio signal s' .

[0052] In Figs. 9A to 9E, a simple implementation of the method for adjusting time information content variations is visualized. By means of an evaluation of the signal energy shown in Fig. 9A, pauses are determined which are illustrated in Fig. 9B in the form of hatched rectangles. The determination of the pauses has located the pauses in time intervals in which the signal energy of the audio signal s is relatively low and possibly close to zero. When the energy is below a certain threshold, the presence of a pause is assumed and hence detected. In addition, security zones are inserted at both ends of the detected pauses in order to prevent a removal of low energy parts of words, such as "F" or "H" sounds. The security zones are represented as thick lines to the left and right of each detected pause in Fig. 9C.

[0053] Fig. 9D illustrates how a ratio of pauses versus speech activity is calculated. The time interval d_1 represents the duration of a first segment containing speech activity (including the security zone). The time interval d_2 represents the duration which is available to the time-stretching function 818 (Fig. 8), when the left pause is used to this end. The time-interval d_2 does not consider that this particular pause may also be utilized by the center segment of speech activity. This may be resolved at a later stage by calculating an average split point within the pause. The calculation of the average split point may be a weighted average calculation based on the individual durations of the various time segments containing speech activity.

[0054] Fig. 9E shows the result after the time-stretching or compression has been performed in accordance with the preparatory calculations.

[0055] It is to be noted that although the duration of the modified audio signal s' shown in Fig. 9E is longer than the duration of the original audio signal s , this is not necessarily the case. In particular, the three segments containing speech activity illustrated in Figs. 9A to 9E, may be maintained at their respective temporal positions, if desired. As such, the time instants of the beginning, the end, or the middle of the segments with each activity may be fixed, and hence, equal in the original audio signal s and in the modified audio signal s' .

[0056] The time-stretching can be done by stretching speech segments into adjacent pauses.

[0057] In the alternative, an estimation of a pause identity can be performed over time, the result of which may then be used for the actual time-stretching or compression. Based on the detected pauses, a speech stretching function is calculated which, among others, limits the variation of the stretching as illustrated in Fig. 10A. Fig. 10A shows a function of single time-stretch factors as a step function. Fig. 10B shows a function of interpolated time manipulation factors or stretching factors based on the step function shown in Fig. 10A. When time-stretching or compressing is based on interpolated time manipulation factors, the listener may more easily adapt to the gradually increasing or decreasing speech velocity as opposed to the abrupt changes of the time manipulation factors shown in Fig. 10A which might lead to equally abrupt changes of the speech velocity in the modified audio signal s' . The interpolation of the time manipulation factors may be performed by a manipulation factor smoother for smoothing the manipulation factor over time.

[0058] Fig. 10C shows a limited variation of the time stretching factors. This fixes the minimal and maximal allowable time-stretching and/or compression. The minimal and maximal threshold needs to be determined, for example, by the fact that excessive time manipulation factors may lead to an unnatural rendering of the audio signals. Furthermore, the sound quality may suffer when a given audio signal or a time section thereof is stretched too much, since the original audio signal only contains a limited number of samples if it is available in a digital format (e.g. PCM). In principle, also an analog signal typically suffers from a loss of sound quality when it is time stretched or shrunk by e.g. electro-mechanical means.

[0059] Fig. 10D also shows a limited variation of the time-stretching / compression, however, adapted to the signal. The degree of time-stretching or compression changes slowly with the signal. The variations within short time segments, however, are limited. The slowly varying lower and upper thresholds $\Delta D_{\min}(t)$ and $\Delta D_{\max}(t)$ may be determined by moving averages over relatively long time intervals, e.g. 10 seconds, 30 seconds, or 1 minute, or values in between.

[0060] Fig. 10E shows the time dilatation function for alternative embodiments of the audio signal processor 100 and the method for adjusting time information content variations. Pauses are not cut or deleted, but remain in the audio signal. Only the regions with speech activity are time-stretched or "compressed", whereas the "filled" pauses remain unmodified.

[0061] The teachings disclosed in this document, in particular, the audio signal processor, the method for adjusting time information content variation and the computer program, enable to time-stretch / compress audio signals in a signal adaptive manner without human interaction. It is possible to detect filled or empty pauses and to process them differently from active speech segments. Moreover, it is possible to playback audio signals slower while maintaining the pitch.

[0062] In particular, spoken language can be played back at a lower speed and thus be more easily comprehensible without necessarily lengthening the duration of the audio signal.

[0063] In the alternative, the total duration may be modified if the pauses are maintained with their original duration. However, the pauses do not need to be time-stretched or compressed along with the rest of the audio signal, so that the new total duration is shorter than a new total duration that would have been obtained by globally time-stretching the entire audio signal. The same applies in principle to the compression of the audio signal so that the total duration of the audio signal, after compression using the proposed methods, would be longer than a total duration of a conventionally (across-the-board) time-compressed audio signal.

[0064] The audio signal processor 100 may further comprise a deleting device implemented to delete the content of the second time section when the second measure of information content M_2 is lower than a deletion threshold. The deletion or erasing of the content in the second time section may be useful if the content comprises repeated word, repeated syllables, makeshift words etc.. Without the deletion, these words, syllables, and sounds would be for example compressed and thus played back at a higher speed than originally recorded, which might be distracting for the listener. In order to identify signal passages in the audio signal s that contain superfluous words or sounds, a pattern detector may be used which compares the signal passages with reference signal passages stored in a database. The reference signal passages may comprise the above mentioned makeshift words when pronounced by various speakers, superfluous sounds such as throat clearing, and other similar patterns. Word repetitions and syllable repetitions may be detected e.g. by an autocorrelation function. Note that word repetitions may be common and perfectly correct in some languages (for example in German) which should be taken into account by a word or syllable repetition function. An excision means may be used to remove the repeated words or syllables from the audio signal.

[0065] The teachings disclosed in this document may be employed in the field of the distribution of audio content, such as digital radio, internet streaming, and audio communication applications. In particular, applications are imaginable in two categories:

- real-time applications, e.g. speech communication and audio coding; and
- processing of already recorded material, e.g. radio plays, lectures, etc.

[0066] The teachings disclosed herein may be beneficial for persons wishing to follow foreign languages more easily or studying foreign languages. Access to radio plays and audiobooks is facilitated to mentally challenged people as well as senior citizens. Furthermore,

applications in the field of training linguistically challenged persons are also possible.

[0067] Some original audio signals may comprise rather long pauses. If these pauses are compressed so that the listener does not need to wait a long time between two speech activity segments, a sound or speech synthesizer may insert a short information about the original duration of the pauses, such as a succession of short beeps, each beep representing a pause of, for example, one minute. A duration of the pause could also be represented by using different pitches of the sound, a low sound indicating the long pause and a high pitched sound indicating a short pause. A speech synthesizer could be used to insert the words "pause of X minutes Y seconds".

[0068] Although some aspects have been described in the context of an apparatus, it is clear that these aspects also represent a description of the corresponding method, where a block or device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus. Some or all of the method steps may be executed by (or using) a hardware apparatus, like for example, a microprocessor, a programmable computer or an electronic circuit. In some embodiments, some one or more of the most important method steps may be executed by such an apparatus.

[0069] Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a Blue-Ray, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

[0070] Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

[0071] Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may for example be stored on a machine readable carrier.

[0072] Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier.

[0073] In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the computer program runs on a computer.

[0074] A further embodiment of the inventive methods is, therefore, a data carrier (or a digital storage medium,

or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein. The data carrier, the digital storage medium or the recorded medium are typically tangible and/or non-transitional.

[0075] A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may for example be configured to be transferred via a data communication connection, for example via the Internet.

[0076] A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.

[0077] A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

[0078] A further embodiment according to the invention comprises an apparatus or a system configured to transfer (for example, electronically or optically) a computer program for performing one of the methods described herein to a receiver. The receiver may, for example, be a computer, a mobile device, a memory device or the like. The apparatus or system may, for example, comprise a file server for transferring the computer program to the receiver.

[0079] In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

[0080] The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

Claims

1. An audio signal processor (100), comprising:

an analysis means (104) implemented to determine a first measure of information content (M_1) of a first time section of an audio signal and a second measure of information content (M_2) of a second time section;

a manipulation factor unit (106) implemented to determine a time manipulation factor (ΔD_1) for

the first time section in dependence on the first measure of information content (M_1) and the second measure of information content (M_2); a time-stretching and compression device (108) implemented to time-stretch or compress the first time section according to the manipulation factor (ΔD_1) and to treat the second time section differently from the first time section.

2. The audio signal processor (100) according to claim 1, further comprising:

a comparator (204) implemented to compare the first measure of information content (M_1) of the first time section to a threshold and to classify the first time section, in dependence on a respective result of the comparison, as a section having a higher measure of information content or as a section having a lower measure of information content; and

a section boundary means (206) implemented to shift boundaries between the sections having a higher measure of information content and the sections having a lower measure of information content into the sections having a lower measure of information content;

wherein the time-stretching and compression device (108) is further implemented to time-stretch or compress a section having a higher measure of information content by a factor corresponding to the shift of the boundaries of the first time section.

3. The audio signal processor according to claim 2, further comprising:

a limiting device (508) for the time-stretching or compression, wherein the limiting device is implemented to determine a current threshold ($\Delta D_{\min}$, $\Delta D_{\max}$) for the time-stretching or compression of the section having higher information content and to limit the time-stretching and compression to the current threshold.

4. The audio signal processor according to claim 3, wherein the limiting device (508) is implemented to evaluate a moving average of the first measure of information content.

5. The audio signal processor according to claim 3 or 4, wherein the limiting device (508) is further implemented to vary the current threshold over the duration of the audio signal (s) in order to adjust sectional variations of the measure of information content (M_1 , M_2).

6. The audio signal processor (100) according to any one of claims 1 to 5, further comprising:

- a pause density estimator implemented to perform pause density estimation over time, a result of which determines a shifting measure for shifting the boundaries.
7. The audio signal processor (100) according to any one of claims 1 to 6, wherein the analysis means (104) is implemented to identify a certain time section as a pause in the audio signal (s) and to set the manipulation factor (M_1 , M_2) for the certain time section to a neutral value so that the certain time section is not time-stretched or compressed.
8. The audio signal processor according to one of claims 1 to 7, further comprising:
- a speech velocity measuring device (602) implemented to determine whether a time section of the audio signal (s) essentially comprises spoken language, and implemented to determine a speech velocity within the time section;
- a threshold setting device (608) connected to the speech velocity measuring device (602) and implemented to determine, based on the determined speech velocity, at least one threshold for the manipulation factor valid for the time section.
9. The audio signal processor (100) according to one of claims 1 to 8, further comprising:
- a deleting device implemented to delete content of the second time section when the second measure of information content is lower than a deletion threshold.
10. The audio signal processor (100) according to one of claims 1 to 9, wherein the time-stretching and compression means comprises at least one of SOLA, WSOLA, PSOLA, PICOLA, TDHS, MPEX or phase vocoder algorithm.
11. The audio signal processor (100) according to one of claims 1 to 10, further comprising:
- a total signal time-stretching and compression means implemented to time-stretch or compress time sections having a higher measure of information content and to leave time sections having a lower measure of information content substantially unaltered with regard to their duration.
12. The audio signal processor (100) according to any one of claims 1 to 11, further comprising:
- a manipulation factor smoother for smoothing the manipulation factor over time.
13. The audio signal processor (100) according to any one of claims 1 to 12, further comprising:
- a repetition detector implemented to detect repeated passages within the audio signal;
- an excision means implemented to excise repeated passages from the audio signal.
14. A method for adjusting time information content variations of an audio signal (s), comprising:
- determining (702) a first measure of information content (M_1) of a first time section of the audio signal and a second measure of information content (M_2) of a second time section of the audio signal (s);
- determining (716) a time manipulation factor (ΔD_1) for the first time section in dependence on the first measure of information content (M_1) and the second measure of information content (M_2);
- processing the audio signal (s) such that the first time section is time-stretched or compressed according to the time manipulation factor (ΔD_1) and that the second time section is processed differently from the first time section.
15. A computer program having a program code for performing the method according to claim 14 when the program runs on a computer.

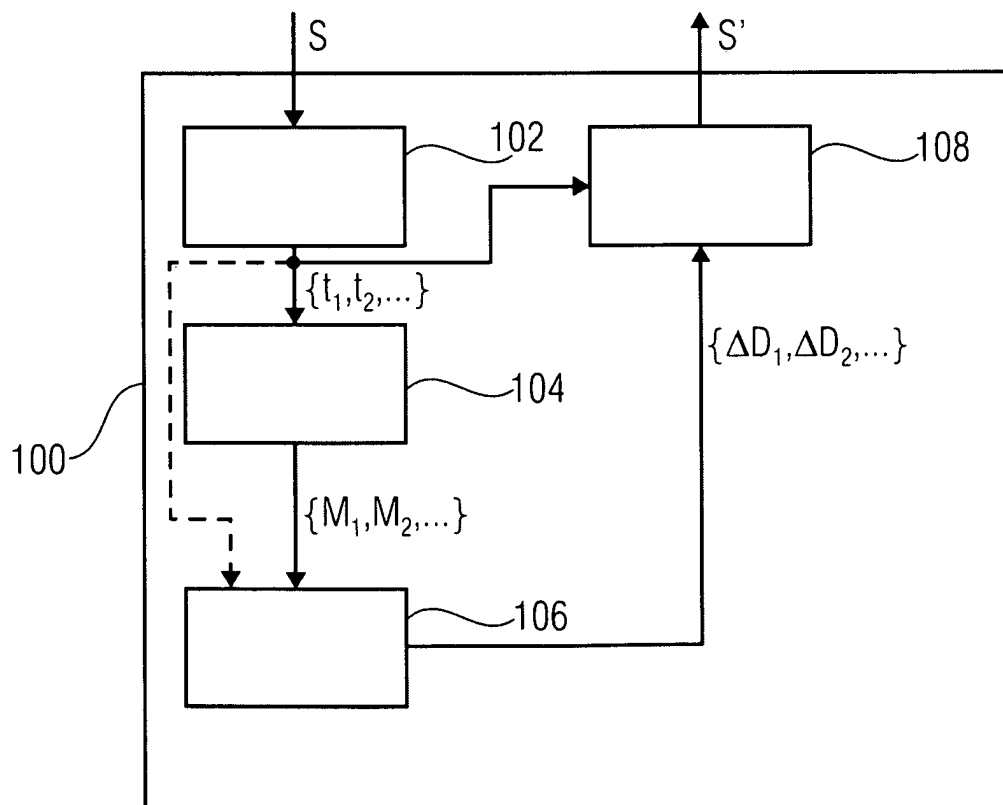
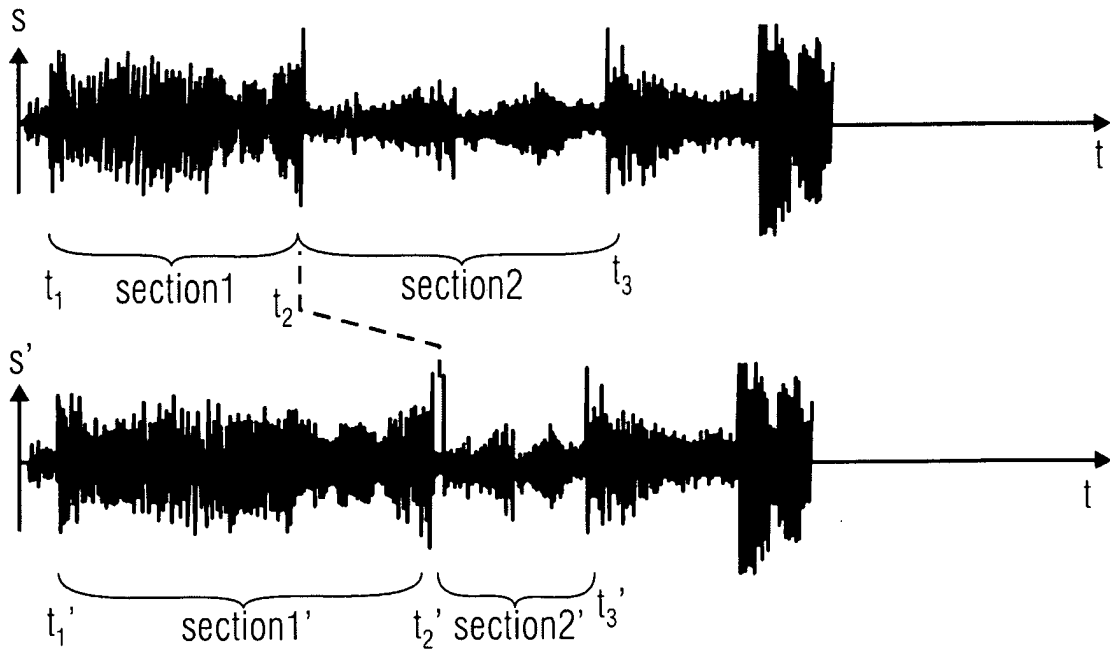


FIG 1

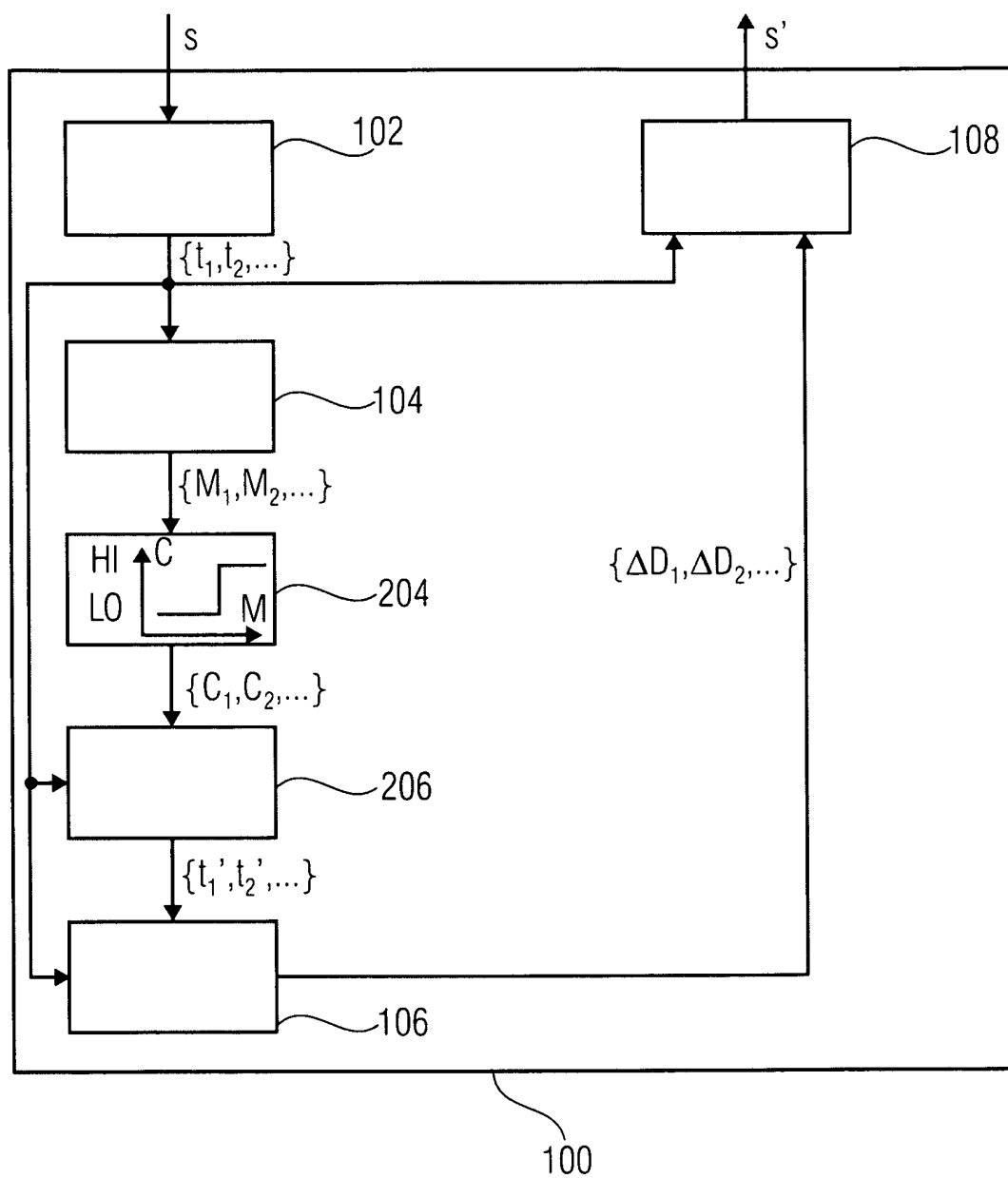


FIG 2

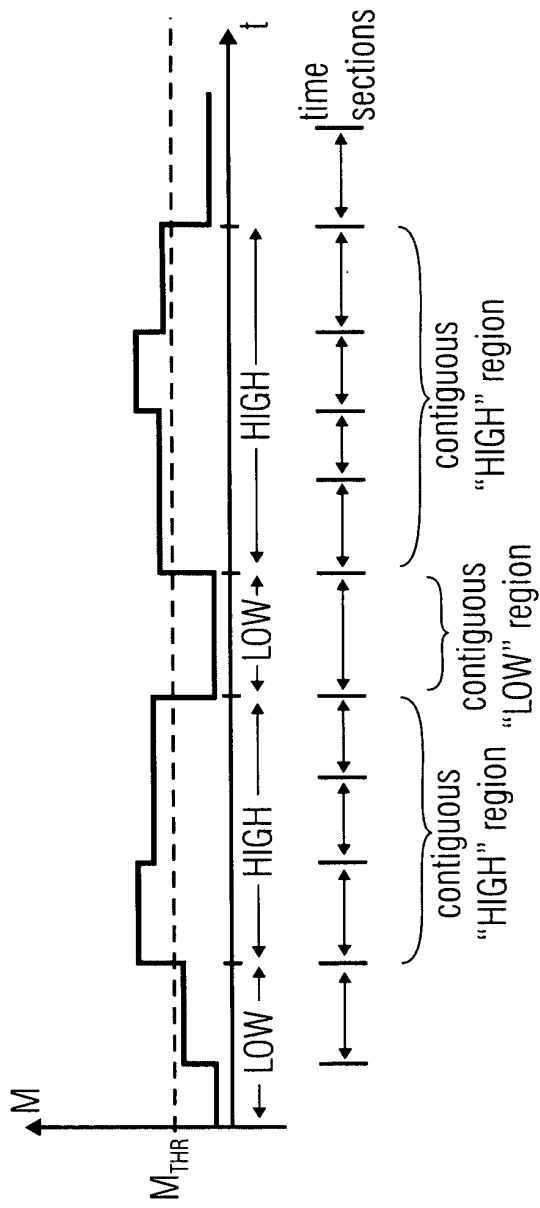


FIG 3

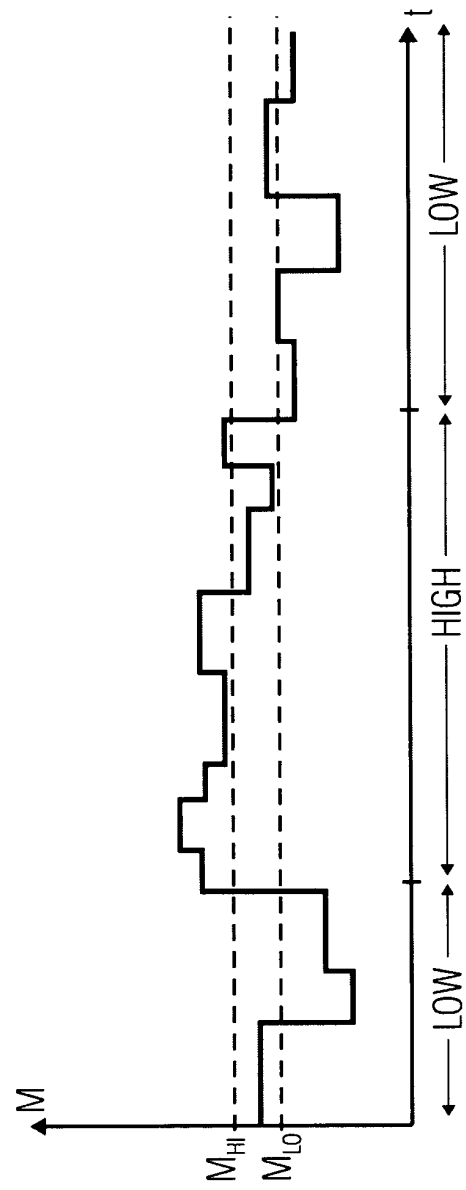


FIG 4

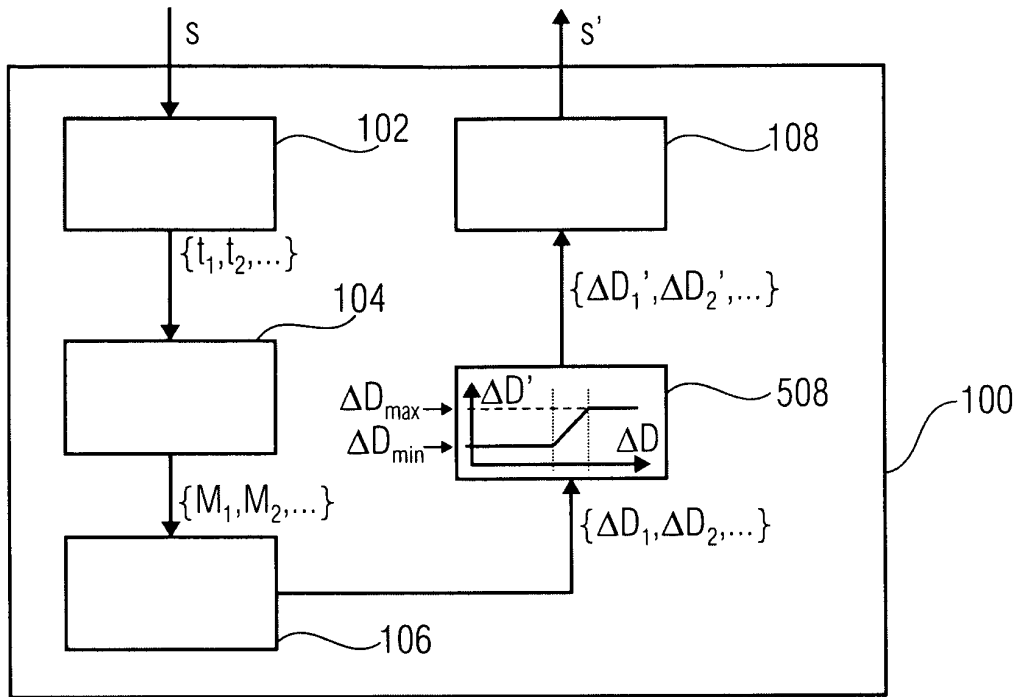


FIG 5

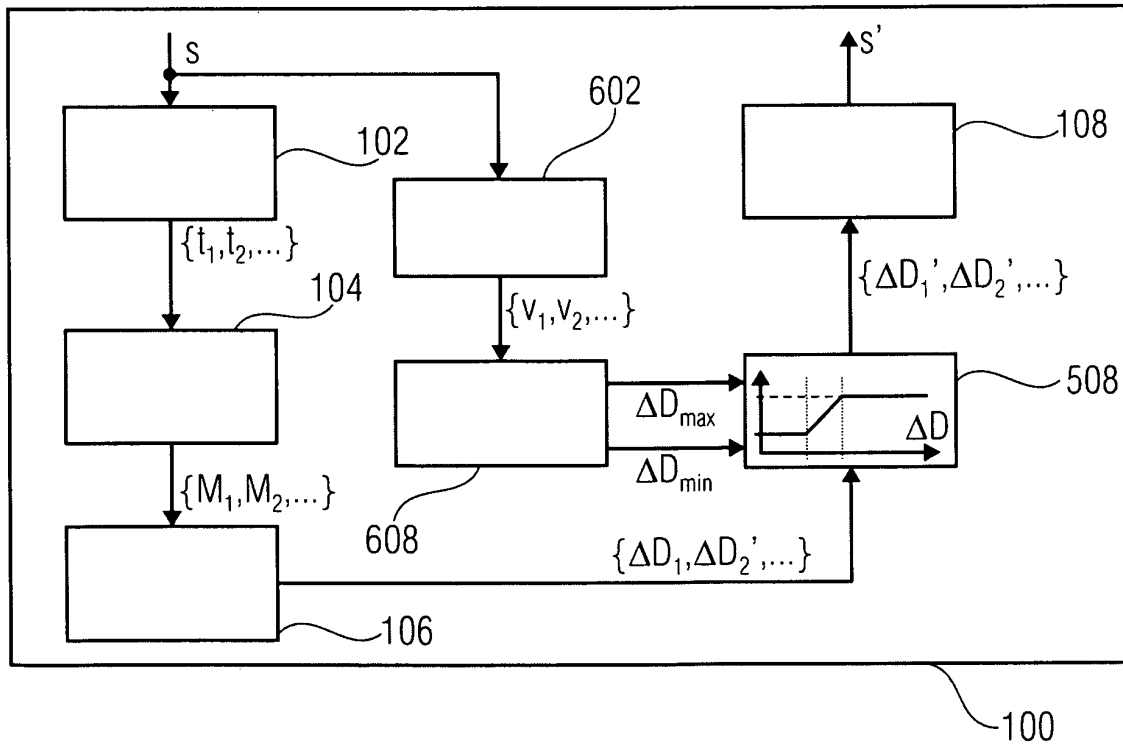


FIG 6

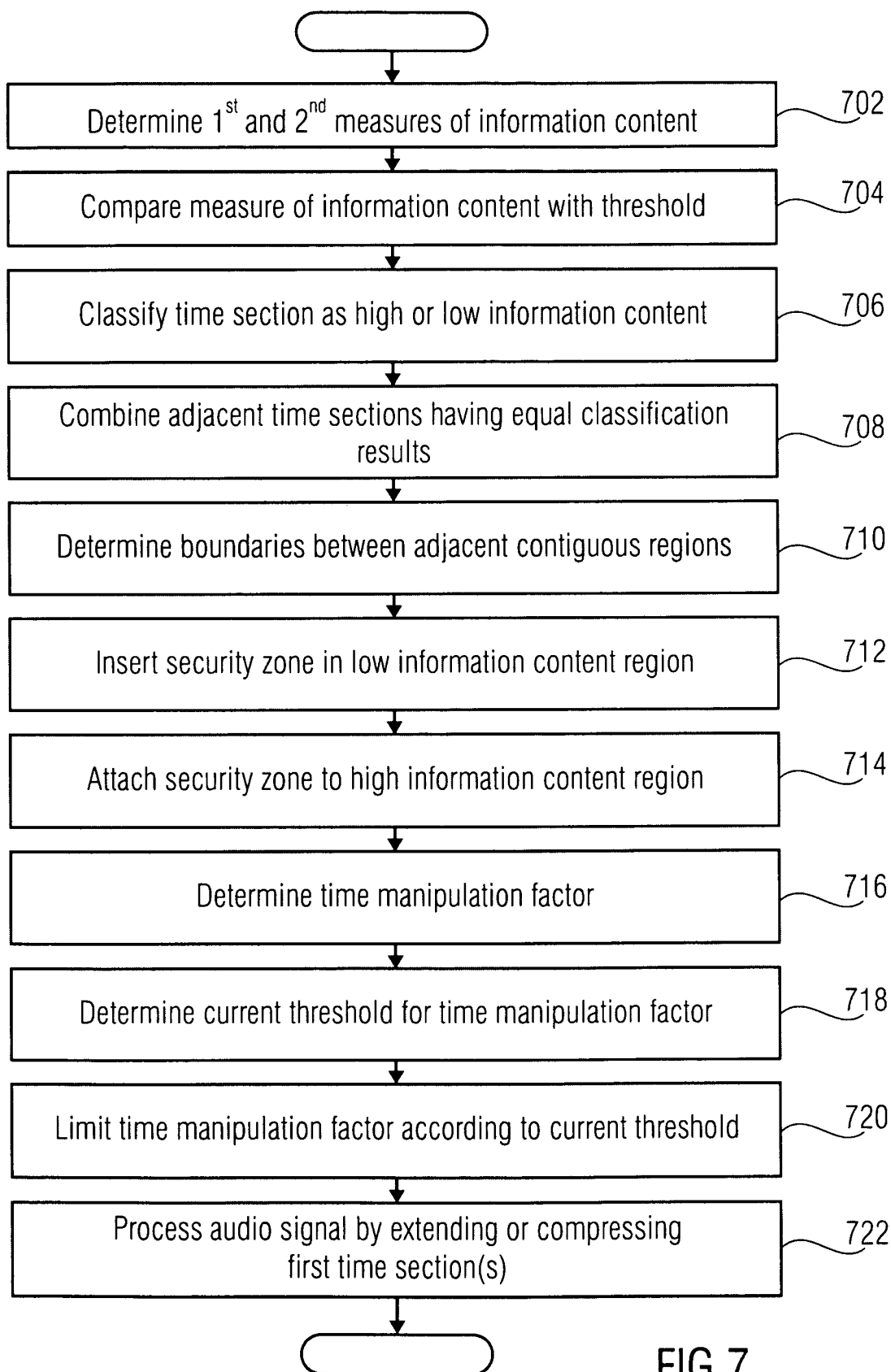


FIG 7

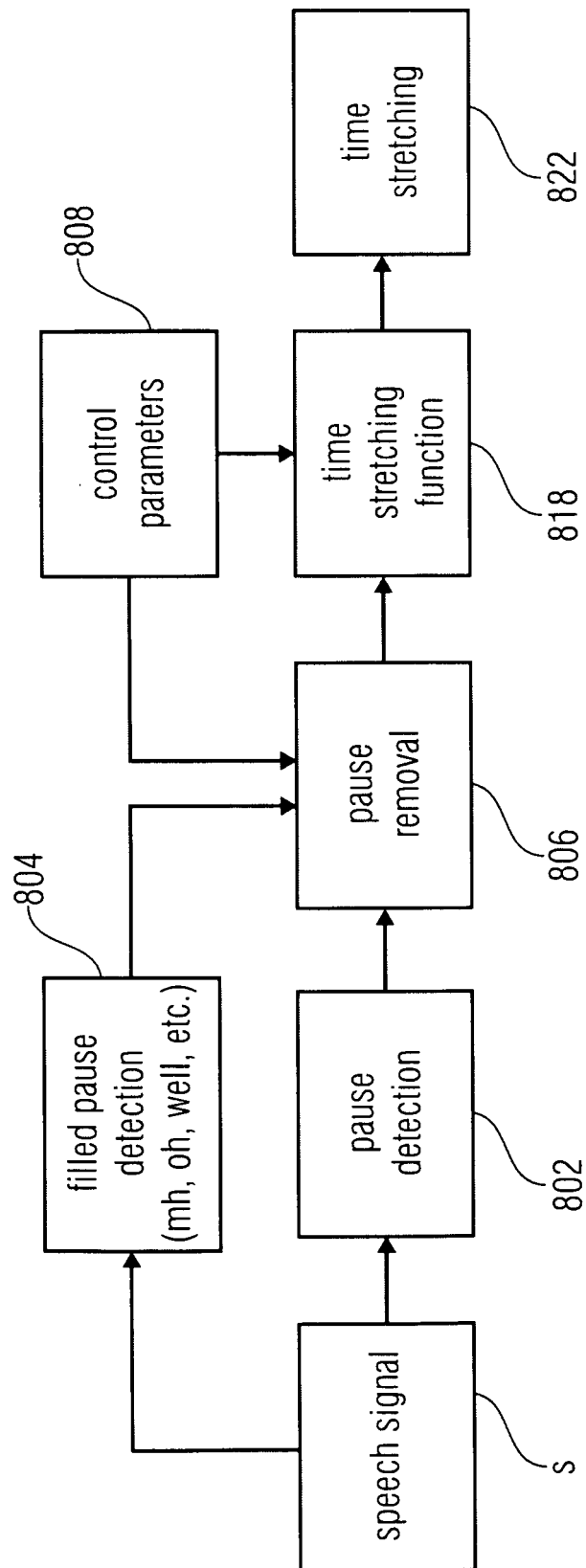
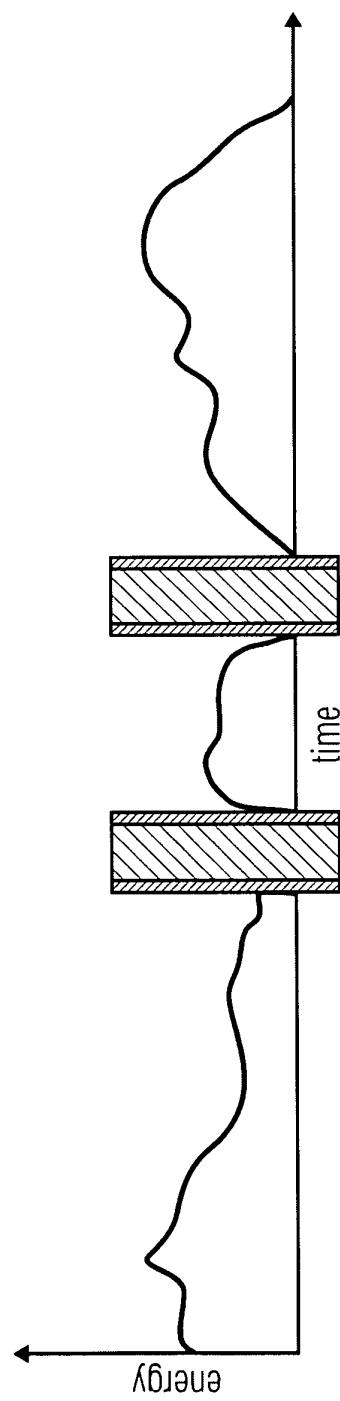
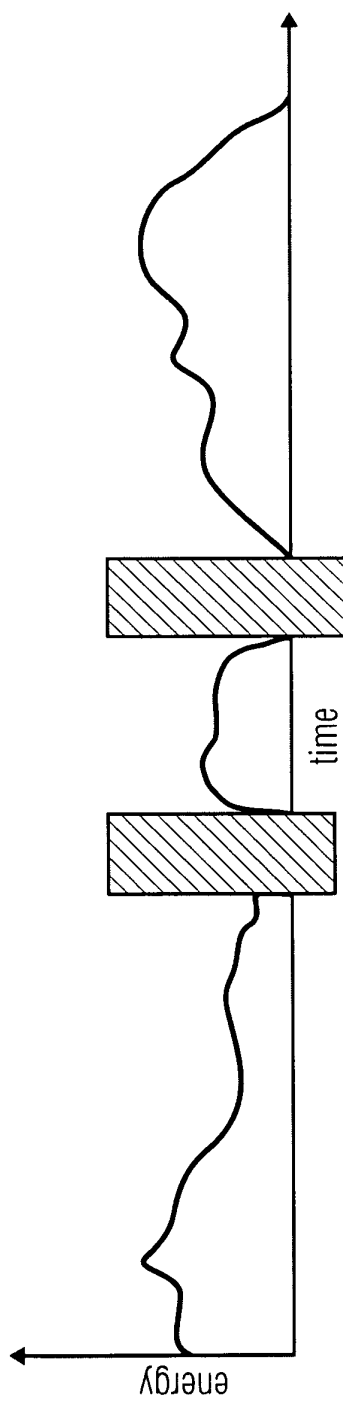
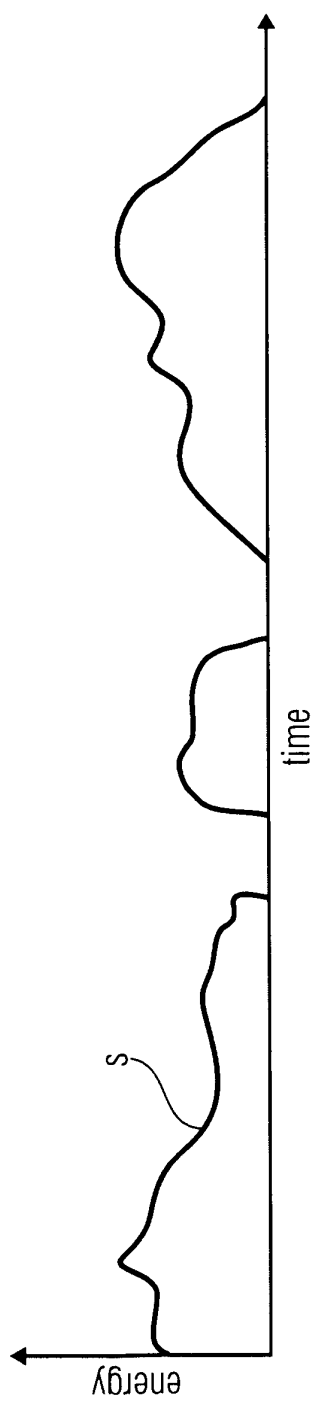


FIG 8



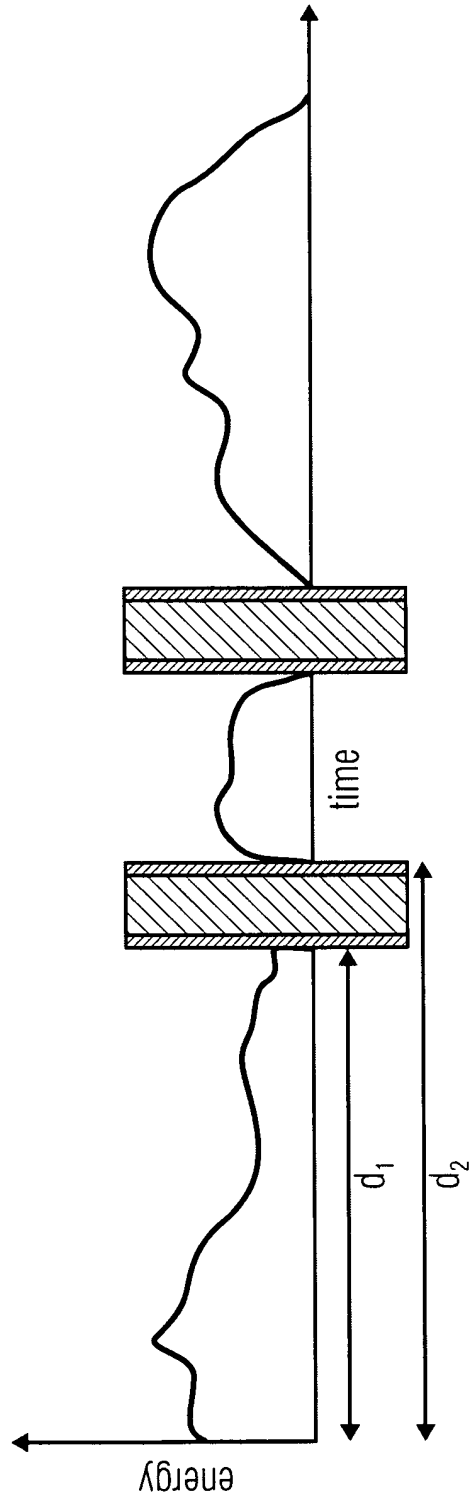


FIG 9D

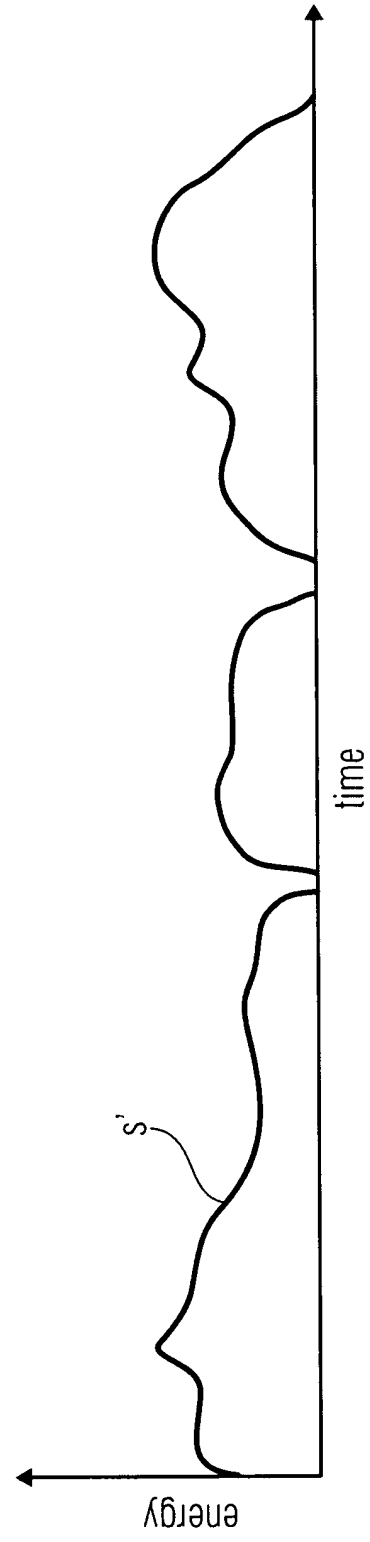


FIG 9E

FIG 10A

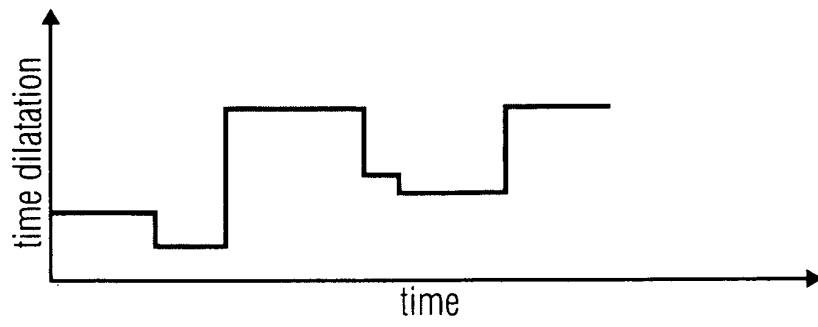


FIG 10B

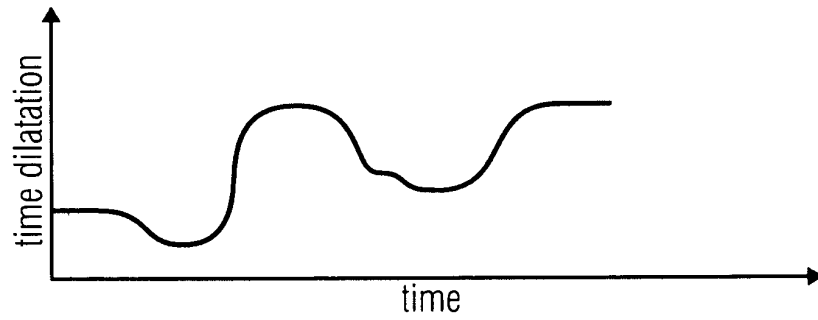


FIG 10C

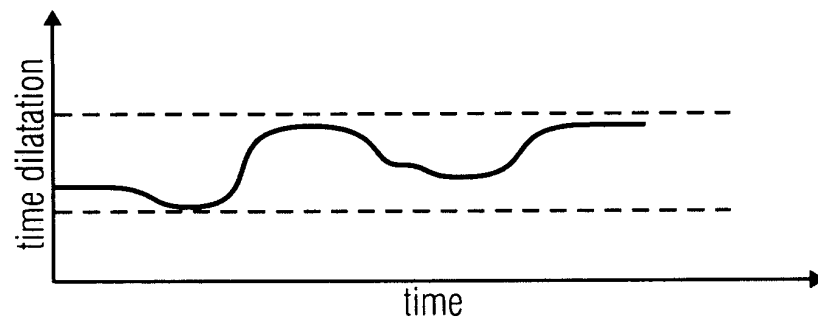


FIG 10D

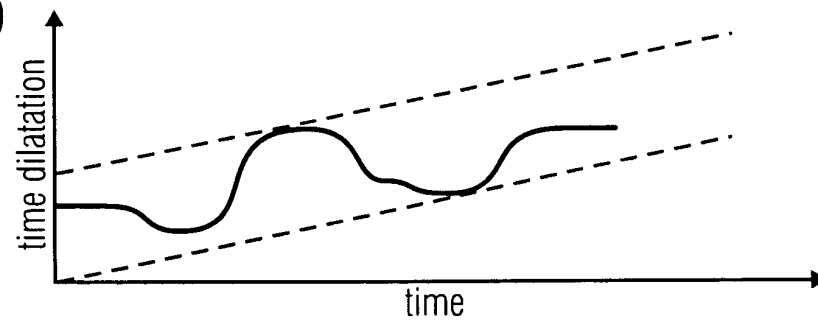
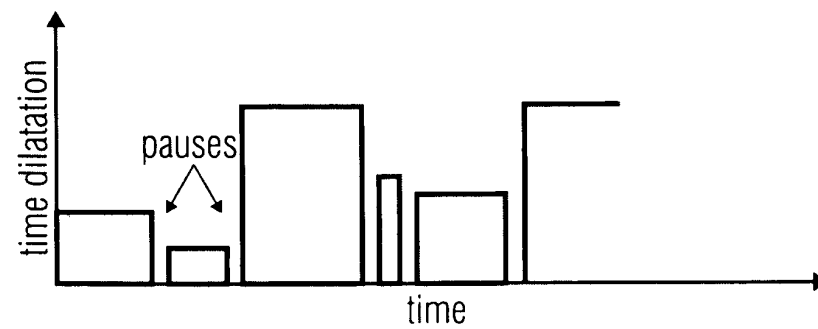


FIG 10E





EUROPEAN SEARCH REPORT

Application Number
EP 11 15 5349

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 5 828 994 A (COVELL MICHELE [US] ET AL) 27 October 1998 (1998-10-27) * abstract * * column 7, line 57 - column 9, line 13 * -----	1-15	INV. G10L21/04
X	EP 1 770 688 A1 (FUJITSU LTD [JP]) 4 April 2007 (2007-04-04) * paragraph [0056] - paragraph [0068]; figures 14-17 * -----	1-15	
			TECHNICAL FIELDS SEARCHED (IPC)
			G10L
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 22 July 2011	Examiner Zimmermann, Elko
CATEGORY OF CITED DOCUMENTS 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 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			

1
EPO FORM 1503 03.82 (P04C01)

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

EP 11 15 5349

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.

22-07-2011

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5828994	A	27-10-1998	AU 719955 B2	18-05-2000
			AU 2829497 A	05-01-1998
			CA 2257298 A1	11-12-1997
			EP 0978119 A1	09-02-2000
			JP 2000511651 A	05-09-2000
			WO 9746999 A1	11-12-1997

EP 1770688	A1	04-04-2007	WO 2006008810 A1	26-01-2006
			JP 4460580 B2	12-05-2010
			US 2007118363 A1	24-05-2007

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

- DE 102008015702 A1 [0007]
- US 6549884 B [0030]

Non-patent literature cited in the description

- **T. KARRER ; E. LEE ; J. BORCHERS.** PhaVoRIT: A Phase Vocoder for Real Time Interactive Time-Stretching. *Proc. ICMC*, November 2006, 708-715 [0003]
- **MARK DOLSON.** The Phase Vocoder: A Tutorial. Computer Audio Research Laboratory [0006]
- **MARK DOLSON.** The Phase-Vocoder: A Tutorial. *Computer Music Journal*, 1986, vol. 10 (4), 14-27 [0030]
- **JEAN LAROCHE ; MARK DOLSON.** New Phase-Vocoder Techniques for Pitch-Shifting, Harmonizing and Other Exotic Effects. *Proceedings 1999 IEEE, Workshop on Applications of Signal Processing to Audio and Acoustics*, 17 October 1999, 91-94 [0030]
- **A. RÖBEL.** New Approach to Transient Processing in the Phase Vocoder. *Proceedings of the International Conference on Digital Audio Effects of DAFX-03*, 08 September 2003, DAFX-1-DAFX6 [0030]
- **MELLER PUCKETTE.** Phase-locked Vocoder. *Proceedings 1995 IEEE, ASSP, Conference on Applications of Signal Processing to Audio and Acoustic Noise*, 1995 [0030]