

(19) World Intellectual Property
Organization
International Bureau



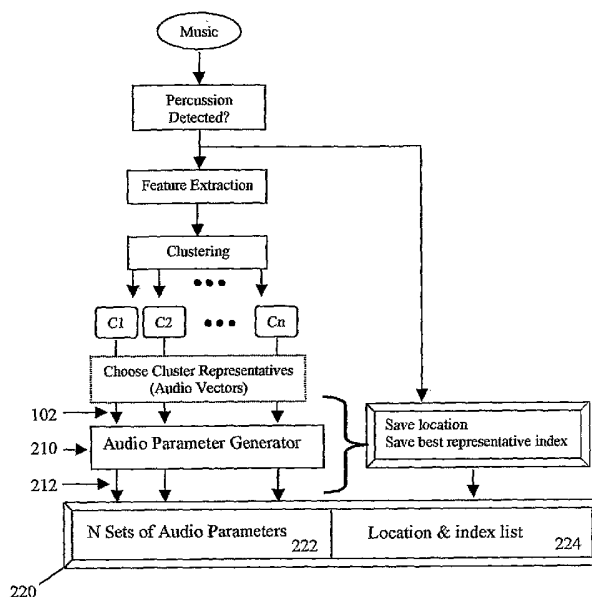
(43) International Publication Date
29 December 2004 (29.12.2004)

PCT

(10) International Publication Number
WO 2004/114134 A1

- (51) International Patent Classification⁷: **G06F 11/00**
- (21) International Application Number:
PCT/SG2004/000187
- (22) International Filing Date: 23 June 2004 (23.06.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/480,783 23 June 2003 (23.06.2003) US
- (71) Applicants (for all designated States except US):
AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH [SG/SG]; 20 Biopolis Way, #07-01 Centros, Singapore 138668 (SG). **NATIONAL UNIVERSITY OF SINGAPORE** [SG/SG]; 10 Kent Ridge Crescent, Singapore 119620 (SG).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **WYSE, Lonce, LaMar** [US/SG]; 1F Pine Grove, #13-31, Singapore 595001 (SG). **WANG, Ye** [FI/SG]; Block C, 111 Clementi Road, #03-06, Singapore 129792 (SG).
- (54) Title: **SYSTEMS AND METHODS FOR CONCEALING PERCUSSIVE TRANSIENT ERRORS IN AUDIO DATA**
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]



(57) Abstract: A method for processing source audio data to conceal percussive transient errors in audio data includes generating one or more of audio vectors from the source audio data, each of the audio vectors representing one or more percussive transients present within the source audio data. Next, a set of audio parameters for each audio vector is defined. Subsequently, the source audio data and the plurality of audio parameter sets are transmitted to the receiving device. If a portion of the source audio data comprising a percussive transient is not received, one or more of the audio parameter sets is used to synthesize a copy of the lost portion of the source audio data.



Published:

— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SYSTEMS AND METHODS FOR CONCEALING PERCUSSIVE TRANSIENT ERRORS IN AUDIO DATA

[0001] The present application relates generally to systems and methods for
5 concealing errors in audio data, and more specifically to systems and methods for
concealing percussive transient errors in audio data.

BACKGROUND

[0002] Communications over the internet and wireless channels are often
10 accomplished using a method of sending a sequence of short packets of data. When used
for audio, the packets represent (possibly overlapping) sequential segments of time in the
audio stream. For CD quality music, each packet typically represents 46 ms (1/20 of a
second) of sound.

[0003] A common kind of error in such systems is lost packets. The baseline
15 standard method for recovering from lost packets is to simply repeat the packet
representing the previous segment of data in the sequence (See e.g. PCT patent
publication no. WO 01/67436 entitled "Sub-Packet Insertion for Packet Loss
Compensation in Voice Over IP Networks" by Bastin; PCT patent publication no. WO
98/13965 entitled "Error Concealment in Digital Audio Receiver," by Sydanmaa et al.;
20 and U.S. Pat. No.5,673,363 entitled "Error Concealment Method and Apparatus of Audio
Signals" by Jeon et al.). There are more sophisticated methods, for example, the missing
data may be extrapolated from already received data (See e.g. U.S. Pat. No. 6,421,802
entitled "Method for Macking Defects in a Stream of Audio Data," by Schildbach et al.).
The extrapolation is "synthetic" in the sense that the exact data used to fill in for the
25 missing packet is not already in the data stream, but no information from the missing
packet is used in the recreation, nor is a model of a specific class of sounds used in the
recreation.

[0004] The repeated packet strategy of error recovery is particularly inappropriate
on or near a musical "beat", the points of strong rhythmic pulse that is characteristic of
30 popular music (referred to herein as a "percussive transient"). If the lost packet occurs
during a steady-state segment of the music, repeating the previous packet is adequate. In
other instances, however, a percussive transient error may occur. For example, if the lost
packet occurs exactly on a percussive transient, such as a musical beat, repeating the

previous packet does not restore the beat, and the violation of the beat expectation makes the error extremely noticeable. Also, if the lost packet immediately follows a beat, the conventional packet replacement strategy creates a very noticeable double percussive transient. Therefore it is desirable to employ a system which can correct these percussive
5 transient errors.

[0005] Fig. 1A illustrates a system 100 operable to correct percussive transient errors, the system disclosed in Wang, Y., Tang, J., Ahmaniemi, A., Vaalgamaa, M., (2003) "Parametric Vector Quantization for Coding Percussive Sounds in Music," ICASSP2003, Hong Kong. In that system, a segment of music is loaded, and a
10 determination made as to whether any percussive transients are detected in the supplied segment. Next, each of the percussive transients are extracted from the segment, and subsequently clustered where each is placed in a particular category along with others having similar audio effects. For example, high-hat beats may be clustered separately from snare drumbeats, which are separately clustered from bass drumbeats. Other
15 percussive transients may be detected and clustered as well.

[0006] Subsequently, a representative vector (herein referred to as an "audio vector") from each cluster group is selected, and the selected audio vectors 102 stored in a header file 120. The header file 120 also includes a codebook index 124 which includes temporal information as to where (i.e., in which frame) a particular beat occurs.
20 The header file 120 is then transmitted in advance of the audio data to a receiving device (not shown) for decoding and recovery.

[0007] Fig. 1B illustrates an exemplary audio stream in which a lost data packet containing a percussive transient is replaced in accordance with the prior art. The receiving device receives the file header 120 in advance of the audio packet stream 130,
25 the file header 120 include the audio vectors 122 and the codebook index 124. From the audio vectors 122 and the codebook index 124, the decoding system is operable to reconstruct the beat patterns throughout the segment. Accordingly, should a beat frame be lost during transmission, a replacement frame 132 is generated and inserted in the packet stream to provide the missing beat or other percussive sound.

30 [0008] While the aforementioned system represents a significant advance over previous work, some disadvantages remain, perhaps the most limiting of which is the amount of bandwidth required to communicate the header file 120. For example, if 8 representative audio vectors are used, and 12 parameters describe the control of the frame

reconstruction model in the receiver, then the aforementioned system requires 32 Kbytes of memory (44.1 KHz sampling rate, 2 bytes per sample, 46ms single-channel vectors, PCM encoded). A greater allocation of bandwidth would be needed to transmit segments of higher fidelity, or segments having more complex beat patterns as a greater number of
5 representative audio vectors would be required.

[0009] What is therefore needed is an improved method for communicating percussive transient information in a bandwidth efficient manner.

SUMMARY OF THE INVENTION

10 [0010] The present invention provides systems and methods for processing audio data to conceal percussive transient errors that enables the data to be communicated in a highly bandwidth efficient manner. The improved efficiency is achieved by generating a set of audio parameters for each of the representative audio vectors, and communicating these parameter sets instead of the entire audio vectors. The audio parameter sets require
15 significantly less bandwidth, and the header file containing this data can be typically reduced in size by 1-2 orders of magnitude compared to the conventional approach in which audio vectors are communicated in the header files.

[0011] The aforementioned features are realized through one method of the present invention in which source audio data is processed for transmission to a receiving
20 device. The method includes generating one or more of audio vectors from the source audio data, each of the audio vectors representing one or more percussive transients present within the source audio data. Next, a set of audio parameters for each audio vector is defined. This audio parameter definition process includes, in one embodiment, processes of computing a parameter set from the original audio vector, synthesizing a
25 audio vector from the defined parameters, comparing the original audio vector to the resynthesized version, and adopting the defined parameter set if the original and synthesized audio vectors are within a predefined range of each other. Subsequently, the source audio data and the plurality of audio parameter sets are transmitted to the receiving device. If a portion of the source audio data comprising a percussive transient
30 is not received at the receiving device, one or more of the audio parameter sets is used to synthesize a copy of the lost packet of the source audio data.

[0012] These and other features of the present invention will be better understood when viewed in light of the following drawings and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] Fig. 1A illustrates a system operable to correct percussive transient errors as known in the art.

5 [0014] Fig. 1B illustrates an audio stream in which a lost data packet containing a percussive transient is replaced in accordance with the prior art.

[0015] Fig. 2A illustrates a method for processing audio data to correct percussive transient errors in accordance with one embodiment of the present invention.

10 [0016] Fig. 2B illustrates a system for processing audio data to correct percussive transient errors in accordance with one embodiment of the present invention.

[0017] Fig. 2C illustrates an exemplary audio stream in which a lost data packet containing a percussive transient is replaced in accordance with one embodiment of the present invention.

15 [0018] Fig. 3A illustrates a system operable to define a parameter set for each of the representative audio vectors in accordance with one embodiment of the present invention.

[0019] Fig. 3B illustrates a method for defining a parameter set for each of the representative audio vectors in accordance with one embodiment of the present invention.

20 [0020] Fig. 4 illustrates a method for computing a set of audio parameters from a representative audio vector in accordance with one embodiment of the present invention.

[0021] Fig. 5A illustrates a contour envelope of a percussive transient modeled in accordance with one embodiment of the present invention.

25 [0022] Fig. 5B illustrates a residual of a percussive transient modeled in accordance with one embodiment of the present invention.

[0023] For clarity, previously-identified features retain their reference numerals in subsequent drawings.

DETAILED DESCRIPTION OF SPECIFIC EMBODIMENTS

30 [0024] Figs. 2A and 2B illustrate a method and system, respectively, for processing audio data to correct percussive transient errors in accordance with one embodiment of the present invention. Referring first to the method of Fig. 2A, initially at 202 the source audio data is supplied to the system 250 and representative audio vectors

are generated therefrom. This process may be performed using those techniques implemented in the conventional system shown in Fig. 1A and prior art.

[0025] Next at 204, a set of audio parameters (212 in Fig. 2B) is defined for each representative audio vector 102, and a codebook index is generated. The defined set of audio parameters 212 represent the audio attributes or characteristics of the percussive transient. For example in one embodiment described below, the audio parameter set includes noisiness, pitch, pitchedness (a combination of the noisiness and pitch parameters), spectral contour, attack duration, and decay duration parameters. These represent only a few of the many possible audio parameters which can be used to describe an audio signal, as those of skill in the art will appreciate, that others may be used alternatively or in addition to these. The audio parameters are computed by means of an audio parameter generator (210 in Fig. 2B), which is further described and illustrated below.

[0026] The generated codebook index describes the frame location of each packet within the source audio stream that contains a percussive transient. This location information is used at the receiver side to identify the correct replacement packet for the frame having a lost packet. The systems and process operable to perform these functions are described in the conventional system of Fig. 1A and in the prior art.

[0027] Lastly at 206, the collection of audio parameter sets 222 and the codebook index 224 are stored in a header file 220, which is preferably transmitted to the receiving device ahead of the source audio data. In one embodiment, the header file and audio data is communicated to the receiving device using the same communication technique or protocol. In an alternative embodiment, a more reliable communication technique or protocol is used to communicate the header file to the receiving device.

[0028] Fig. 2C illustrates an exemplary audio stream in which a lost audio packet containing a percussive transient is replaced in accordance with one embodiment of the present invention. The received audio data includes a file header 220 and an output audio packet stream 230, with the file header 220 including the collection of audio parameter sets 222 and a codebook index 224 generated by the system 250 shown in Fig. 2B. From the collection of parameter sets 222, a collection of audio vectors 226 are synthesized, and from the collection of synthesized audio vectors, a collection of packet frames 228 of the output audio stream 230 containing percussive transients are reconstructed.

[0029] Once the audio packets containing percussive transients are reconstructed, the received audio packet stream 230 is assembled for playback. A lost packet file in the output audio stream 230 is detected using the means implemented in the conventional system of Fig. 1A and prior art. Where the lost packet frame contains a percussive transient, the codebook index 224 identifies the appropriate reconstructed packet to be inserted in the target frame. The identified packet is inserted and the audio output stream, the processes of which are described in the conventional system of Fig. 1 A and prior art.

[0030] Figs. 3A and 3B illustrate the audio parameter generator 210 and method of operation, respectively, in accordance with embodiments of the present invention. The generator 210 includes an analysis and parameter extraction module 304, an audio vector synthesizer 306, feature extraction modules 308, and a comparator 310. The implementation of each of these modules may be realized in software, firmware, hardware or a combination of these.

[0031] During operation, a representative audio vector 102 ("original audio vector") is supplied to the analysis and extraction module 304 (process 351, Fig. 3B). Module 304 is operable to examine the original audio vector, and compute therefrom a set of audio parameters 212 which models the vector's audio attributes (process 353). A number of parameters may be used to model the audio vector, some examples being noisiness, pitch, spectral contour, attack duration, and decay duration. Other audio parameters, a few examples being Spectral centroid, Mel Frequency Cepstral Coefficients (MFCC's) or actual drum types (snare + bassdrum), may be used alternatively or in addition to these.

[0032] The computed parameter set 212 is subsequently supplied to the audio vector synthesizer 306, which generates a synthesized audio vector 309 therefrom (process 355). The synthesizer 306 may comprise any conventionally known synthesizer operable to generate the audio vector responsive to the receipt of the audio parameter set.

[0033] Feature extraction modules 308 are operable to extract one or more features of the original and synthesized audio vectors, features in the illustrated exemplary embodiment being LPC coefficients, pitch, pitchedness, peak time and level. The reader will appreciate that other features may be extracted as well, e.g., bass drummedness, as well as other sounds or parameters thereof.

[0034] The extracted features of the original and synthesized audio vectors 102 and 309 are supplied to a comparator 310 (in one embodiment, a summer having one negative input terminal as shown), and the difference obtained. If the difference is outside of the predefined range, the parameter set 212 is re-computed, and the processes of 353-355 are repeated. Once the extracted features of the original and synthesized vectors are within the predefined range, the presently computed parameter set is adopted as defining the supplied audio vector 102. The predefined range will depend upon the particular extracted feature. For example, in the instance in which the extracted feature is pitchedness, the predefined range may be [0,1], where 0 denotes unpitched noise and 1 denotes that substantially all energy is concentrated at harmonics of a fundamental frequency of the pitch). In another embodiment in which the extracted feature is spectral centroid the predefined range may be [0, 10,000 Hz]. The reader will appreciate that it is possible to employ other extraction features and values in alternative embodiments under the present invention.

[0035] Fig. 4 illustrates a method for computing the set of audio parameters from a representative audio vector as shown in Fig. 3B in accordance with one embodiment of process 353 shown in Fig. 3B. Initially at 402, the contour envelope of the percussive transient is modeled. In a particular embodiment of this process, an onset slope and a decay slope are used to model the transient. The maximum point of the signal is identified and used as the vertex of the formed contour triangle.

[0036] Next at 404, the overall spectral shape of audio vector is modeled using a first group of audio parameters. In a particular embodiment, a twelve coefficient linear predictive analysis (LPC) is used to compute the overall spectral shape, each of the twelve coefficients comprising a parameter of the audio parameter set 212. Of course, a larger or smaller number of coefficients may be used to compute the vector spectrum, depending upon the desired trade-off between spectral accuracy -v- header file size, as each of the LPC coefficients comprises an audio parameter stored within the header file 220.

[0037] Next at 406, the residual shape of the audio vector is computed using a second set of audio parameters. This process, in a particular embodiment, is achieved by modeling the residual of the above LPC computation as a pitched signal plus white noise. In a specific embodiment, the second group of parameters used in computing the residual shape of the audio vector include amplitude peak time, amplitude peak level, pitch, and

pitchedness. Amplitude peak time is obtained by computing the short-time averaged root mean square (rms) value and finding the location of the maximum, and amplitude peak level is computed by computing the rms value at the peak time. These parameters are functionally equivalent to the aforementioned parameters of attack duration and decay duration.

[0038] Pitch is computed by determining the maximal peak of the FFT-derived power spectrum autocorrelation within the frequency range of 100-500 HZ. Pitchedness is computed as a ratio of the peak to total power in the spectrum. The gain of the residual may be used as a part of the vector synthesis process in 306.

[0039] In one embodiment of the present invention, the codebook index is included within the header file and communicated to the receiving device. In an alternative embodiment, an audio parameter set for individual percussive transients are sent to the receiving device in the packet previous to the percussive transient, the parameter set operable to regenerate a lost packet in real time on the receiver side. In such an embodiment, the codebook index data could be omitted from the header file.

[0040] Figs. 5A and 5B illustrates exemplary embodiment of the event contour and residual of the LPC processes described in Fig. 4 above. As shown, the vectors are modeled as a combination of noise and periodic information with a single broad spectral shape and a fixed duration of 2,048 PCM samples. Sixteen audio parameters are used to model the vectors, twelve LPC coefficients, and the amplitude peak time, amplitude peak level, pitch, and pitchedness parameters described above. Of course, a different number of parameters (e.g., 32 LPC coefficients), and/or different parameter types (e.g. Mel Frequency Cepstral Coefficients (MFCC's)) can be used in alternative embodiments under the present invention.

[0041] Within the receiving device, the audio above-described audio parameters are used to generate a copy of the original audio data. Initially, the residual is synthesized, which in one embodiment, includes using white noise as the source and applying a comb filter with a delay corresponding to the pitch. The pitchedness parameter is used to determine the filter weights – the relative balance between the delay tap and white noise. The pitch parameter is then used to amplitude modulate the noisy signal with a sharp attack and exponential decay at the pitch period, while the pitchedness parameter is used to control the decay rate - a longer decay rate makes the amplitude modulation less pronounced and the signal less pitched. Subsequently, the

course spectral shape is recovered using the LPC-derived filter. The temporal amplitude contour is then applied to the regenerated signal to recover the audio vector. The audio data can then be generated from the recovered audio vector using the techniques employed in the prior art system of Fig. 1A. The codebook index maps the receiver-generated audio packets to frames within the output audio stream, such that missing audio packets containing a percussive transient can be replaced using the receiver generated packets.

[0043] In comparison to the prior art device shown in Fig. 1A, the present invention provides a savings in header file size on the order of 1-2 degrees of magnitude. Assuming for both systems that 8 representative audio vectors are used, and that 12 parameters describe the control of the audio vector synthesizer 306, the prior art system will accordingly require approximately 32 kbytes of memory (44.1kHz sampling rate, 2 bytes per sample, 46ms single-channel vectors, PCM encoded). In comparison, the present invention requires less than 200 bytes. In those embodiments in which the synthesis model is transmitted with the data (for example, when the decoder does not possess the synthesis model, or the user has developed a custom synthesis model on the encoder side and needs to convey it to the decoder side for proper sound reproduction) the additional header room needed is on the order of 2-6 kbytes, which still represents a significant decrease in the header file size required in the conventional system.

[0044] As readily appreciated by those skilled in the art, the described processes may be implemented in hardware, software, firmware or a combination of these implementations as appropriate. For example, the processes of defining an audio parameter set for a given audio vector may be carried out as software instruction code, whereas the processes of transmitting signals may be achieved using hardware electronics. Further, some or all of the described processes may be implemented as computer readable instruction code resident on a computer readable medium (removable disk, volatile or non-volatile memory, embedded processors, etc.), the instruction code operable to program a computer or other such programmable device to carry out the intended functions.

INCORPORATED REFERENCES

[0045] The following references are herein incorporated by reference in their entirety for all purposes:

Wang, Y., Tang, J., Ahmaniemi, A., Vaalgamaa, M., (2003) "Parametric Vector Quantization for Coding Percussive Sounds in Music," ICASSP2003, Hong Kong;

S. Levine and J.O. Smith, (1999) "A switched parametric & transform audio coder," Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Phoenix, 1999;

5 Scheirer, E. D., (2001) "Structured Audio, Kolmogorov Complexity, and Generalized Audio Coding," IEEE Transactions on Speech and Audio Processing, Vol.9, No. 8, November 2001; and

Karplus, K., and A. Strong. 1983. "Digital Synthesis of Plucked-String and Drum Timbres". Computer Music Journal 7(2):43-55. Reprinted in C. Roads, ed. 1989. The
10 Music Machine. Cambridge, Massachusetts: MIT Press.

[0046] While the above is a detailed description of the present invention, it is only exemplary and various modifications, alterations and equivalents may be employed in various aspects and processes described herein. Accordingly, the scope of the present invention is hereby defined by the metes and bounds of the following claims.

15

CLAIMSWhat is claimed is:

1. A method for processing source audio data to conceal percussive transient errors,
5 the method comprising:
generating one or more of audio vectors from the source audio data, each of the
audio vectors representing one or more percussive transients present within the source
audio data;
defining a set of audio parameters for each audio vector; and
10 transmitting the source audio data and the plurality of audio parameter sets to a
receiving device,
wherein, if a portion of the source audio data comprising a percussive transient is
not received at the receiving device, one or more of the audio parameter sets is used to
synthesized a copy of the lost portion of the source audio data.
15
2. The method of claim 1, wherein defining a set of audio parameters for each audio
vector comprises:
receiving an original audio vector
computing a set of audio parameters from each original audio vector;
20 synthesizing a synthetic audio vector from the set of computed audio parameters;
and
comparing the synthetic and original audio vectors;
wherein if the synthetic and original audio vectors are within a predefined range,
the computed set of audio parameters are defined as the set of audio parameters for the
25 particular audio vector.
3. The method of claim 1, further comprising generating a codebook index
which describes the frame location of each packet within the source audio stream that
contains a percussive transient; and wherein transmitting further comprises transmitting
30 the codebook index to the receiving device.

4. The method of claim 3, further comprising creating a header file, the header file comprising the plurality of parameter sets and the codebook index, and wherein transmitting comprises transmitting the header file ahead of the source audio data.

5 5. The method of claim 2, wherein computing a set of audio parameters from each audio vector comprises:

modeling the contour envelope of the percussive transient;

modeling the spectral shape of the audio vector using a linear predictive analysis based upon a first predefined group of audio parameters; and

10 modeling the residual error signal based upon a second predefined group of audio parameter,

wherein the first and second predefined groups of audio parameters collective comprise the set of audio parameters.

15 6. The method of claim 5, wherein the second predefined group of audio parameters comprises: (i) an amplitude peak time audio parameter, (ii) an amplitude peak level parameter, (iii) a pitch parameter, and (iv) a pitchedness parameter.

7. A system for processing source audio data to conceal percussive transient errors,
20 the system comprising:

means for generating one or more of audio vectors from the source audio data, each of the audio vectors representing one or more percussive transients present within the source audio data;

means for defining a set of audio parameters for each audio vector; and

25 means for transmitting the source audio data and the plurality of audio parameter sets to a receiving device,

wherein, if a portion of the source audio data comprising a percussive transient is not received at the receiving device, one or more of the audio parameter sets is used to synthesized a copy of the lost portion of the source audio data.

30

8. The system of claim 7, wherein the means for defining a set of audio parameters for each audio vector comprises:

means for computing a set of audio parameters from each original audio vector;

means for synthesizing a synthetic audio vector from the set of computed audio parameters; and

means for comparing the synthetic and original audio vectors;

wherein if the synthetic and original audio vectors are within a predefined range,
5 the computed set of audio parameters are defined as the set of audio parameters for the particular audio vector.

9. The system of claim 7, further comprising means for generating a codebook index which describes the frame location of each packet within the source audio stream that
10 contains a percussive transient; and wherein the means for transmitting further comprises means for transmitting the codebook index to the receiving device.

10. The system of claim 9, further comprising a means for creating a header file, the header file comprising the plurality of parameter sets and the codebook index, and
15 wherein the means for transmitting comprises a means for transmitting the header file ahead of the source audio data.

11. The system of claim 8, wherein the means for computing a set of audio parameters from each audio vector comprises:

20 means for modeling the contour envelope of the percussive transient;

means for modeling the spectral shape of the audio vector using a linear predictive analysis based upon a first predefined group of audio parameters; and

means for modeling the residual error signal based upon a second predefined group of audio parameter.

25 wherein the first and second predefined groups of audio parameters collective comprise the set of audio parameters.

12. The system of claim 11, wherein the second predefined group of audio parameters comprises: (i) an amplitude peak time audio parameter, (ii) an amplitude peak level
30 parameter, (iii) a pitch parameter, and (iv) a pitchedness parameter.

13. A method for processing a received sequence of audio data into an audio stream, the received audio data including source audio data and a plurality of audio parameter

sets, each of the audio parameter sets defining one or more percussive transients present within the source audio data, the method comprising:

assembling the received sequence of audio data into an audio stream;

determining if the assembled audio stream omits a portion of the audio stream
5 comprising a percussive transient; and

generating a copy of the missing portion of the assembled audio stream using the one or more of the audio parameter sets if the audio stream omits a portion comprising a percussive transient.

10 14. The method of claim 13, wherein the plurality of audio parameter sets is received ahead of the source audio data.

15. The method of claim 13, wherein generating a copy of the missing portion of the assembled audio stream comprises:

15 synthesizing, from the plurality of audio parameter sets, a respective plurality of audio vectors, wherein each audio vector defines one or more percussive transients in the source audio data; and

generating, using one or more of the synthesized audio vectors, a copy of the missing portion of the audio stream.

20

16. The method of claim 15, wherein the received audio data further includes a codebook index operable to map one of the audio packets reconstructed from the set of audio parameters to a particular packet frame, and wherein generating a copy of the missing portion of the assembled audio stream further comprises accessing the codebook
25 index to determine that the generated copy of the missing packet is mapped to the packet frame comprising the missing portion of the assembled audio stream.

17. A system for processing a received sequence of audio data into an audio stream, the received audio data including source audio data and a plurality of audio parameter
30 sets, each of the audio parameter sets defining one or more percussive transients present within the source audio data, the system comprising:

means for assembling the received sequence of audio data into an audio stream;

means for determining if the assembled audio stream omits a portion of the audio stream comprising a percussive transient; and

means for generating a copy of the missing portion of the assembled audio stream using the one or more of the audio parameter sets if the audio stream omits a portion
5 comprising a percussive transient.

18. The system of claim 17, wherein the plurality of audio parameter sets is received ahead of the source audio data.

10 19. The system of claim 17, wherein the means for generating a copy of the missing portion of the assembled audio stream comprises:

means for synthesizing, from the plurality of audio parameter sets, a respective plurality of audio vectors, wherein each audio vector defines one or more percussive transients in the source audio data; and

15 means for generating, using one or more of the synthesized audio vectors, a copy of the missing portion of the audio stream.

20 20. The system of claim 19, wherein the received audio data further includes a codebook index operable to map one of the audio packets reconstructed from the set of audio parameters to a particular packet frame, and wherein the means for generating a copy of the missing portion of the assembled audio stream further comprises means for accessing the codebook index to determine that the generated copy of the missing packet is mapped to the packet frame comprising the missing portion of the assembled audio stream.

25

21. A computer program product, resident of a computer-readable medium, which is operable to execute instruction code for processing source audio data to conceal percussive transient errors, the computer program product comprising:

30 instruction code to generate one or more of audio vectors from the source audio data, each of the audio vectors representing one or more percussive transients present within the source audio data;

instruction code to define a set of audio parameters for each audio vector; and

instruction code to transmit the source audio data and the plurality of audio parameter sets to a receiving device,

wherein, if a portion of the source audio data comprising a percussive transient is not received at the receiving device, one or more of the audio parameter sets is used to
5 synthesized a copy of the lost portion of the source audio data.

22. The computer program product of claim 21, wherein the instruction code to defining a set of audio parameters for each audio vector comprises:

instruction code to receive an original audio vector
10 instruction code to compute a set of audio parameters from each original audio vector;

instruction code to synthesize a synthetic audio vector from the set of computed audio parameters; and

instruction code to compare the synthetic and original audio vectors;
15 wherein if the synthetic and original audio vectors are within a predefined range, the computed set of audio parameters are defined as the set of audio parameters for the particular audio vector.

23. The computer program product of claim 21, further comprising instruction code
20 to generate a codebook index which describes the frame location of each packet within the source audio stream that contains a percussive transient; and wherein the instruction code to transmit further comprises instruction code to transmit the codebook index to the receiving device.

24. The computer program product of claim 23, further comprising instruction code
25 to create a header file, the header file comprising the plurality of parameter sets and the codebook index, and wherein the instruction code to transmit comprises instruction code to transmit the header file ahead of the source audio data.

25. The computer program product of claim 22, wherein the instruction code to
30 compute a set of audio parameters from each audio vector comprises:

instruction code to model the contour envelope of the percussive transient;

instruction code to model the spectral shape of the audio vector using a linear predictive analysis based upon a first predefined group of audio parameters; and

instruction code to model the residual error signal based upon a second predefined group of audio parameter,

5 wherein the first and second predefined groups of audio parameters collective comprise the set of audio parameters.

26. The computer program product of claim 25, wherein the second predefined group of audio parameters comprises: (i) an amplitude peak time audio parameter, (ii) an
10 amplitude peak level parameter, (iii) a pitch parameter, and (iv) a pitchedness parameter.

27. A computer program product, resident of a computer-readable medium, which is operable to execute instruction code for processing a received sequence of audio data into an audio stream, the received audio data including source audio data and a plurality of
15 audio parameter sets, each of the audio parameter sets defining one or more percussive transients present within the source audio data, the computer program product comprising:

instruction code to assemble the received sequence of audio data into an audio stream;

20 instruction code to determine if the assembled audio stream omits a portion of the audio stream comprising a percussive transient; and

instruction code to generate a copy of the corrupted or missing portion of the assembled audio stream using the one or more of the audio parameter sets if the audio stream omits a portion comprising a percussive transient.

25

28. The computer program product of claim 27, wherein the plurality of audio parameter sets is received ahead of the source audio data.

29. The computer program product of claim 27, wherein the instruction code to
30 generate a copy of the missing portion of the assembled audio stream comprises:

instruction code to synthesize, from the plurality of audio parameter sets, a respective plurality of audio vectors, wherein each audio vector defines one or more percussive transients in the source audio data; and

instruction code to generate, using one or more of the synthesized audio vectors, a copy of the missing portion of the audio stream.

30. The computer program product of claim 29, wherein the received audio data
5 further includes a codebook index operable to map one of the audio packets reconstructed
from the set of audio parameters to a particular packet frame, and wherein the instruction
code to generate a copy of the missing portion of the assembled audio stream further
comprises instruction code to access the codebook index to determine that the generated
copy of the missing packet is mapped to the packet frame comprising the missing portion
10 of the assembled audio stream.

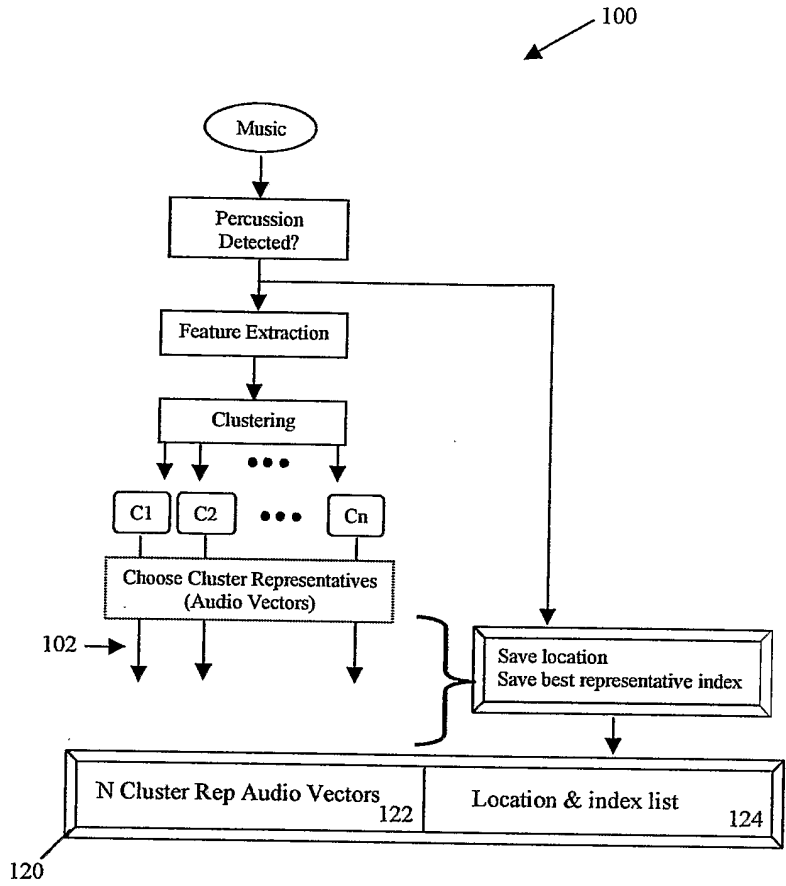


Fig. 1A

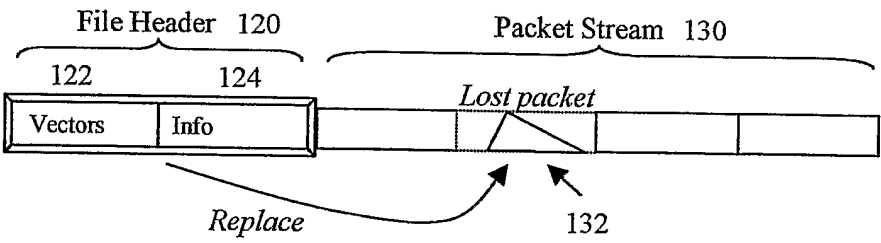


Fig. 1B

2/7

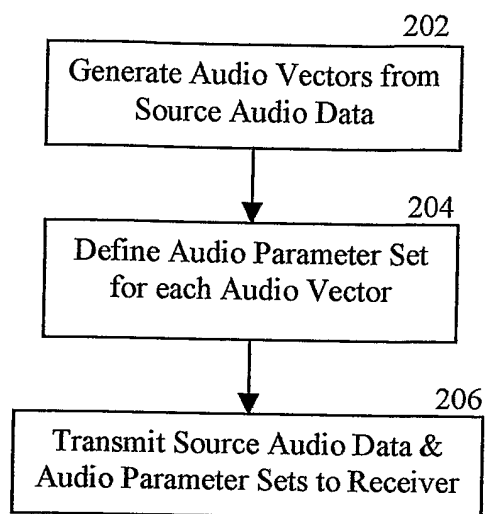


Fig. 2A

3/7

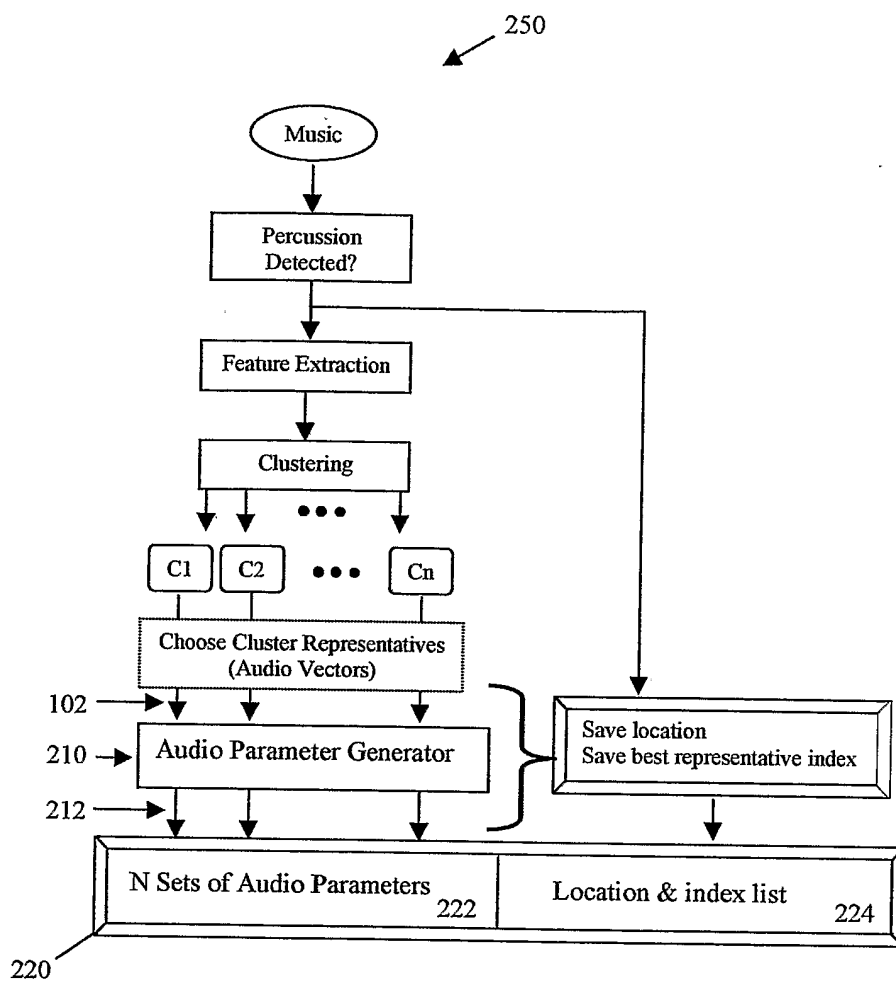


Fig. 2B

4/7

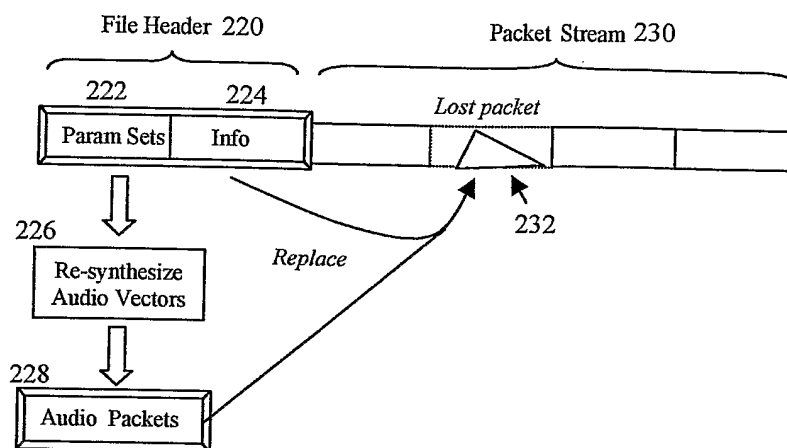


Fig. 2C

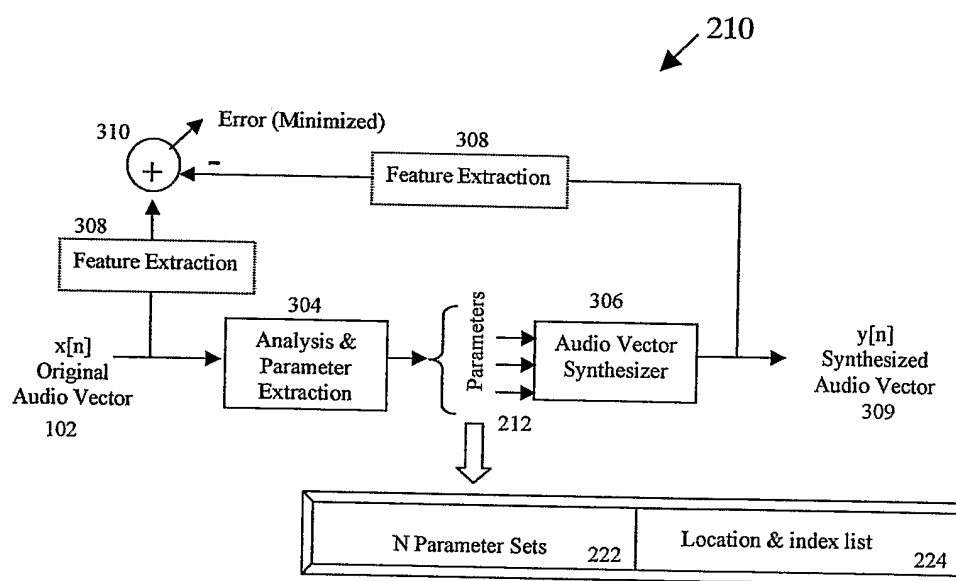


Fig. 3A

5/7

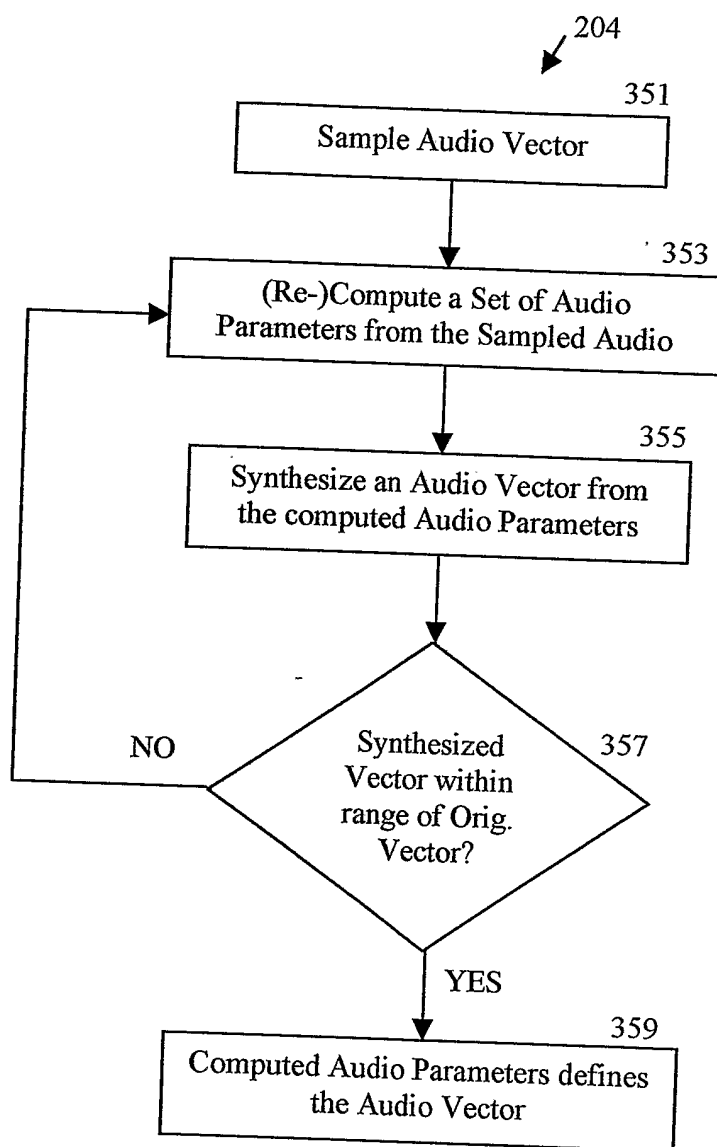


Fig. 3B

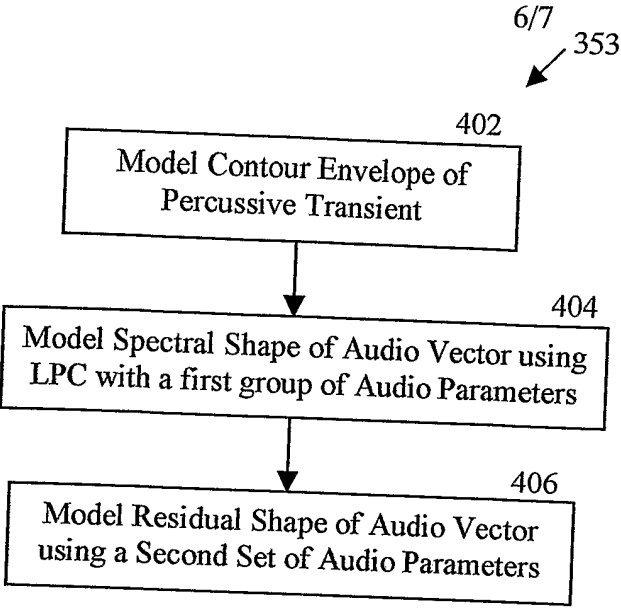


Fig. 4

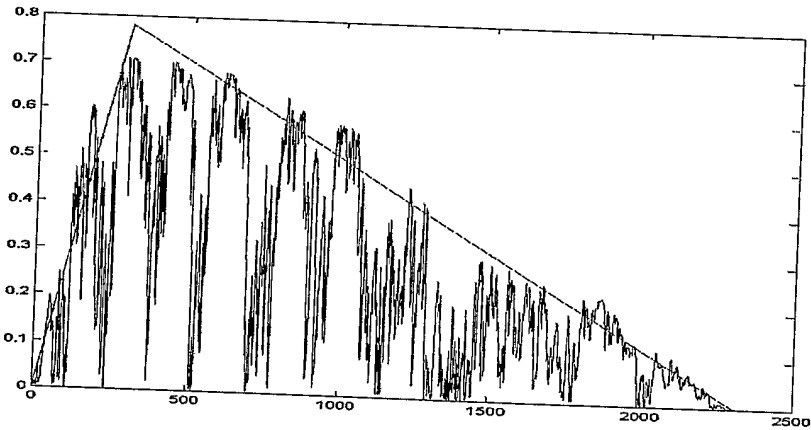


Fig. 5A

7/7

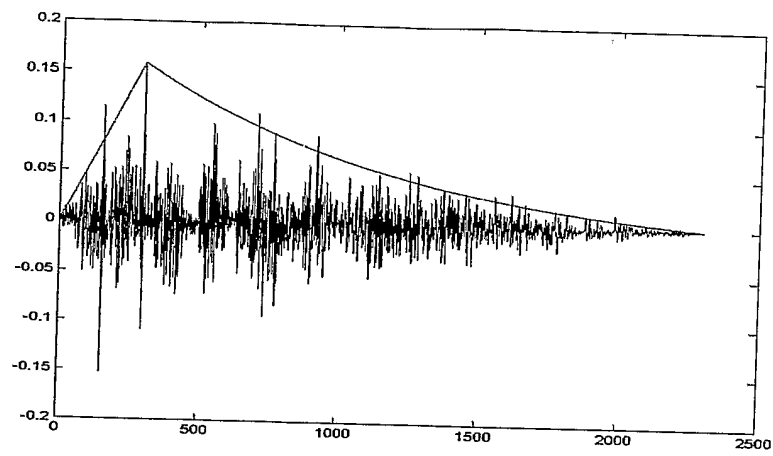


Fig. 5B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2004/000187

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. ⁷: G06F 11/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT, ESP@CE, USPTO, INTERNET, Keywords (audio, beat, transient, data, error, loss, vector, parameter, synthesis) and similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	US 2004/0083110 A1 (WANG) 29 April 2004. See whole document, in particular paragraphs 10 – 20, 61 – 65 and 77 – 83.	1, 7; 13, 14, 17, 18, 21, 27, 28
X	WANG Y., et al., "PARAMETRIC VECTOR QUANTIZATION FOR CODING PERCUSSIVE SOUNDS IN MUSIC", 2003 IEEE International Conference on Acoustics, Speech and Signal Processing, Proceedings (ICASSP 2003), 6 – 10 April 2003, Volume 5, pages 652 – 655. See whole document.	1, 7, 13, 14, 17, 18, 21, 27, 28
A	US 2002/0133764 A1 (WANG) 19 September 2002. See whole document.	

☐ Further documents are listed in the continuation of Box C ☒ See patent family annex

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
23 September 2004

Date of mailing of the international search report
29 SEP 2004

Name and mailing address of the ISA/AU
AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaaustralia.gov.au
Facsimile No. (02) 6285 3929

Authorized officer
BEN TUOHY
Telephone No : (02) 6283 7918

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2004/000187

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in
Search Report

Patent Family Member

US 2004083110 WO 2004038927

US 2002133764 US 2002138795 US 2002178012 WO 02059875

WO 02060070

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX