

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 September 2011 (01.09.2011)

(10) International Publication Number
WO 2011/104283 A1

- (51) International Patent Classification:
G10L 19/00 (2006.01)
- (21) International Application Number:
PCT/EP2011/052694
- (22) International Filing Date:
23 February 2011 (23.02.2011)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
10154948.3 26 February 2010 (26.02.2010) EP
- (71) Applicant (for all designated States except US):
FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. [DE/DE]; Hansastraße 27c, 80686 München (DE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **ZITZMANN, Reinhard** [DE/DE]; Johann-Sebastian-Bach-Str. 25, 91083 Baiersdorf (DE). **WABNIK, Stefan** [DE/DE]; Hultschiner Str. 16a, 26129 Oldenburg (DE). **PICKEL, Joerg** [DE/DE]; Herbsttal 2, 91230 Happurg (DE). **GREEVEN-BOSCH, Bert** [NL/NL]; Roestmos 12, NL-3069 AR Rotterdam (NL). **GRILL, Bernhard** [DE/DE]; Peter-Henlein-Str. 7, 91027 Lauf (DE). **EBERLEIN, Ernst**

[DE/DE]; Waldstrasse 28, 91091 Grosseenseebach (DE). **DEL GALDO, Giovanni** [IT/DE]; Ziegelweg 6a, 90562 Heroldsberg (DE). **KRAEGELOH, Stefan** [DE/DE]; Schobertweg 51, 91056 Erlangen (DE). **BLIEM, Tobias** [DE/DE]; Tennenloher Str. 26, 91058 Erlangen (DE). **BORSUM, Juliane** [DE/DE]; Täublingstraße 22, 91058 Erlangen (DE). **BREILING, Marco** [DE/DE]; Christian-Ernst-Str. 8, 91052 Erlangen (DE).

(74) Agents: **BURGER, Markus** et al.; Postfach 246, 82043 Pullach (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

[Continued on next page]

(54) Title: WATERMARK SIGNAL PROVIDER AND METHOD FOR PROVIDING A WATERMARK SIGNAL

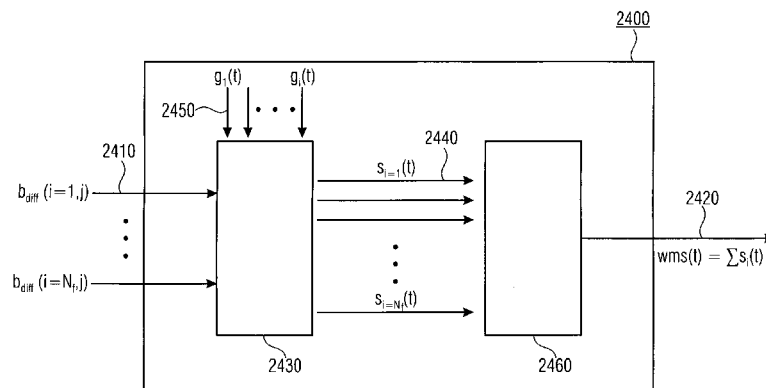
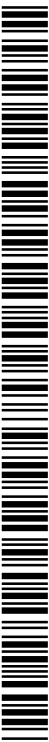


FIGURE 24

(57) Abstract: A watermark signal provider for providing a watermark signal in dependence on a time frequency-domain representation of watermark data, in which the time-frequency-domain representation comprises values associated to frequency subbands and bit intervals, the watermark signal provider comprises a time-frequency-domain waveform provider to provide time-domain waveforms for a plurality of frequency subbands, based on the time-frequency-domain representation of the watermark data. The time-frequency-domain waveform provider is configured to map a given value of the time-frequency-domain representation onto a bit shaping function. A temporal extension of the bit shaping function is longer than the bit interval associated to the given value of the time-frequency-domain representation, such that there is a temporal overlap between bit shaped functions provided for temporally subsequent values of the time-frequency-domain representation of the same frequency subband. A time-domain waveform of a given frequency subband contains a plurality of bit shaped functions provided for temporally subsequent values of the time-frequency-domain representation of the same frequency band. The watermark signal provider further comprises a time-domain waveform combiner, to combine the provided time-domain waveforms for the plurality of frequencies of the time-frequency-domain provider to derive the watermark signal.



WO 2011/104283 A1

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, **Published:**
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, — *with international search report (Art. 21(3))*
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Watermark Signal Provider and Method for Providing a Watermark Signal

Description

5

Technical Field

10 Embodiments according to the present invention are related to a watermark signal provider for providing a watermark signal in dependence on a time-frequency domain representation of watermark data. Further embodiments are related to a method for providing a watermark signal in dependence on a time-frequency domain representation of watermark data.

15 Some embodiments according to the invention are related to a robust low complexity audio watermarking system.

Background of the Invention

20 In many technical applications, it is desired to include an extra information into an information or signal representing useful data or "main data" like, for example, an audio signal, a video signal, graphics, a measurement quantity and so on. In many cases, it is desired to include the extra information such that the extra information is bound to the main data (for example, audio data, video data, still image data, measurement data, text
25 data, and so on) in a way that it is not perceivable by a user of said data. Also, in some cases it is desirable to include the extra data such that the extra data are not easily removable from the main data (e.g. audio data, video data, still image data, measurement data, and so on).

30 This is particularly true in applications in which it is desirable to implement a digital rights management. However, it is sometimes simply desired to add substantially unperceivable side information to the useful data. For example, in some cases it is desirable to add side information to audio data, such that the side information provides an information about the source of the audio data, the content of the audio data, rights related to the audio data and
35 so on.

For embedding extra data into useful data or "main data", a concept called "watermarking" may be used. Watermarking concepts have been discussed in the literature for many

different kinds of useful data, like audio data, still image data, video data, text data, and so on.

In the following, some references will be given in which watermarking concepts are discussed. However, the reader's attention is also drawn to the wide field of textbook literature and publications related to the watermarking for further details.

DE 196 40 814 C2 describes a coding method for introducing a non-audible data signal into an audio signal and a method for decoding a data signal, which is included in an audio signal in a non-audible form. The coding method for introducing a non-audible data signal into an audio signal comprises converting the audio signal into the spectral domain. The coding method also comprises determining the masking threshold of the audio signal and the provision of a pseudo noise signal. The coding method also comprises providing the data signal and multiplying the pseudo noise signal with the data signal, in order to obtain a frequency-spread data signal. The coding method also comprises weighting the spread data signal with the masking threshold and overlapping the audio signal and the weighted data signal.

In addition, WO 93/07689 describes a method and apparatus for automatically identifying a program broadcast by a radio station or by a television channel, or recorded on a medium, by adding an inaudible encoded message to the sound signal of the program, the message identifying the broadcasting channel or station, the program and/or the exact date. In an embodiment discussed in said document, the sound signal is transmitted via an analog-to-digital converter to a data processor enabling frequency components to be split up, and enabling the energy in some of the frequency components to be altered in a predetermined manner to form an encoded identification message. The output from the data processor is connected by a digital-to-analog converter to an audio output for broadcasting or recording the sound signal. In another embodiment discussed in said document, an analog bandpass is employed to separate a band of frequencies from the sound signal so that energy in the separated band may be thus altered to encode the sound signal.

US 5, 450,490 describes apparatus and methods for including a code having at least one code frequency component in an audio signal. The abilities of various frequency components in the audio signal to mask the code frequency component to human hearing are evaluated and based on these evaluations an amplitude is assigned to the code frequency component. Methods and apparatus for detecting a code in an encoded audio signal are also described. A code frequency component in the encoded audio signal is

detected based on an expected code amplitude or on a noise amplitude within a range of audio frequencies including the frequency of the code component.

5 WO 94/11989 describes a method and apparatus for encoding/decoding broadcast or recorded segments and monitoring audience exposure thereto. Methods and apparatus for encoding and decoding information in broadcasts or recorded segment signals are described. In an embodiment described in the document, an audience monitoring system encodes identification information in the audio signal portion of a broadcast or a recorded segment using spread spectrum encoding. The monitoring device receives an acoustically reproduced version of the broadcast or recorded signal via a microphone, decodes the identification information from the audio signal portion despite significant ambient noise and stores this information, automatically providing a diary for the audience member, which is later uploaded to a centralized facility. A separate monitoring device decodes additional information from the broadcast signal, which is matched with the audience diary information at the central facility. This monitor may simultaneously send data to the centralized facility using a dial-up telephone line, and receives data from the centralized facility through a signal encoded using a spread spectrum technique and modulated with a broadcast signal from a third party.

20 WO 95/27349 describes apparatus and methods for including codes in audio signals and decoding. An apparatus and methods for including a code having at least one code frequency component in an audio signal are described. The abilities of various frequency components in the audio signal to mask the code frequency component to human hearing are evaluated, and based on these evaluations, an amplitude is assigned to the code frequency components. Methods and apparatus for detecting a code in an encoded audio signal are also described. A code frequency component in the encoded audio signal is detected based on an expected code amplitude or on a noise amplitude within a range of audio frequencies including the frequency of the code component.

30 However, in the known watermarking systems, a watermark signal is based on a plurality of time domain adjacent waveforms, wherein a maximum energy of this waveforms is limited, because the watermark signal has to be kept inaudible. But a low energy of the waveform and therefore of the watermark signal leads to a more difficult detection of the watermark signal and may lead to bit errors and therefore a low robustness of the water
35 mark signal.

In view of the situation, it is the object of the present invention to create a concept for providing a watermark signal, which allows for an easier decoding of the watermark signal at a receiver side.

5 Summary of the Invention

The objective is achieved by a watermark signal provider according to claim 1, a method for providing a watermark signal according to claim 10 and a computer program according to claim 11.

10

An embodiment according to the present invention creates a watermark signal provider for providing a watermark signal in dependence on a time-frequency domain representation of watermark data. The time-frequency domain representation comprises values associated to frequency subbands and bit intervals. The watermark signal provider comprises a time-
15 frequency domain waveform provider and a time domain waveform combiner. The time-frequency domain waveform provider is configured to map a given value of the time-frequency domain representation onto a bit shaping function. A temporal extension of the bit shaping function is longer than the bit interval associated to the given value of the time-frequency domain representation, such that there is a temporal overlap between bit shaped
20 functions provided for temporally subsequent values of the time-frequency domain representation of the same frequency subband. The time-frequency domain waveform provider is further configured such that a time domain waveform of a given frequency subband contains a plurality of bit shaped functions provided for temporally subsequent values of the time-frequency domain representation of the same frequency band. The time
25 domain waveform combiner is configured to combine the provided waveforms for the plurality of frequencies of the time-frequency domain waveform provider to derive the watermark signal.

It is a key idea of the present invention, to not only correlate binary values (e.g. binary
30 values of the same frequency subband and of subsequent bit intervals) of a representation of watermark data, but also to correlate the bit shaped functions corresponding to this values with each other. In this way a redundancy in the water marked signal is added, which allows for an easier decoding at a receiver side, without raising the energy of the watermark signal. Furthermore a robustness of the watermark signal is increased.

35

This correlation of the bit shaped function is achieved in embodiments by bit shaping functions, wherein a temporal extension of the bit shaping functions is longer than a bit time of corresponding values of the time-frequency domain representation.

Therefore a decoder for the watermark signal at a receiver side can be made easier and less complex than a decoder for a conventional water marking system. Furthermore a chance of obtaining a correct watermark information out of an obtained signal can be increased especially in noisy environments.

Values of the time-frequency domain representation of watermark data may be binary values, wherein one value corresponds to a frequency subband and a bit interval.

10 In an embodiment the time-frequency domain waveform provider is configured to provide a bit shaped function for each of the values of the time-frequency domain representation, wherein the time-frequency domain waveform provider is configured such that bit shaped functions of adjacent values of the same frequency band overlap and therefore a correlation of bit shaped functions of adjacent values is achieved.

15

In an embodiment the time-frequency domain waveform provider may be configured such that a bit shaped function provided for a given value of the time-frequency domain representation is overlapped with a bit shaped function of a temporally preceding value of the same frequency subband like the given value of the time-frequency domain representation and with a bit shaped function of a temporally following value of the same frequency subband like the given value of the time-frequency domain representation, such that a time domain waveform provided by the time-frequency domain waveform provider contains an overlap between at least three temporally subsequent bit shaped functions of the same frequency subband. In other words a time domain waveform of a given frequency subband is in a given bit interval at least based on a first bit shaped function of a first value corresponding to the given frequency subband and the given time interval, on a second bit shaped function of a second value corresponding to the given frequency subband and a temporally preceding time interval and on a third bit shaped function of a third value corresponding to the given frequency subband and a temporally following time interval.

20

30

In an embodiment a temporal extension of a bit shaping function may be a temporal range, in which the bit shaping function comprises non zero values. Furthermore the temporal range, in where the bit shaping function comprises non zero values may be at least three bit intervals long

35

A bit shaping function may also be called a bit forming function and may be different for each frequency subband of the time-frequency domain representation of the watermark

data. Therefore achieving a different filtering (bit shaping) for different frequency subbands.

5 In an embodiment a bit shaping function may be based on an amplitude modulated periodic signal. An amplitude modulation of the amplitude modulated periodic signal may be based on a baseband function. A temporal extension of the bit shaping function may be based on the baseband function. Therefore a temporal extension of the baseband function, wherein the baseband function contains not zero values, is longer than the bit interval. The baseband function may be identical for values of a same frequency band of the time-
10 frequency domain representation of the watermark data.

In an embodiment the baseband function is identical for a plurality or for all of the frequency subbands of the time-frequency domain representation. In other words the baseband function may be the same for a plurality of values or all values of the time-
15 frequency domain representation. If the baseband function is identical for every subband, a more efficient implementation at a decoder side is possible.

In an embodiment an amplitude modulation factor of a bit shaping function may be a time domain baseband function, for example like a filter function. The baseband function may
20 be identical for values of a same frequency band of the time-frequency domain representation of the watermark data.

In an embodiment a periodic part of a bit shaping function of a given frequency subband may be based on a cosinus function, based on a frequency which is a center frequency of
25 the given frequency subband.

In an embodiment the watermark signal provider further comprises a weight tuner, for example a psychoacoustical processing module, which is configured to tune a weight (and therefore an amplitude) of each bit shaped function for each value of the time domain
30 representation of the watermark data. The weight tuner may be configured to maximize an energy of a bit shaped function of a given value in regard of inaudibility of the watermark signal. In other words, the weight tuner may be configured to fine tune the weights to assign as much energy as possible to the watermark while keeping it inaudible.

35 In an embodiment the weight tuner may be configured to tune the weights in an iterative process controlled by the weight tuner. The weight tuner can therefore adjust each bit shaped function provided from the time-frequency domain waveform provider such that

each bit shaped function has a maximum energy (but of course stays inaudible) and therefore is better to detect at a decoder side.

5 In an embodiment a time domain waveform of a given frequency subband is a sum of all bit shaped functions of the given frequency subband .

In an embodiment the watermark signal is a sum of the provided waveforms for the plurality of frequency subbands.

10 Some embodiments according to the invention also create a method for providing a watermark signal in dependence on a time-frequency domain representation of watermark data. That method is based on the same findings as the apparatus discussed before.

15 Some embodiments according to the invention comprise a computer program for performing the inventive method.

Brief Description of the Figures

20 Embodiments according to the invention will subsequently be described taking reference to the enclosed figures, in which:

Fig. 1 shows a block schematic diagram of a watermark inserter according to an embodiment of the invention;

25 Fig. 2 shows a block-schematic diagram of a watermark decoder, according to an embodiment of the invention;

Fig. 3 shows a detailed block-schematic diagram of a watermark generator, according to an embodiment of the invention;

30

Fig. 4 shows a detailed block-schematic diagram of a modulator, for use in an embodiment of the invention;

35 Fig. 5 shows a detailed block-schematic diagram of a psychoacoustical processing module, for use in an embodiment of the invention;

Fig. 6 shows a block-schematic diagram of a psychoacoustical model processor, for use in an embodiment of the invention;

- Fig. 7 shows a graphical representation of a power spectrum of an audio signal output by block 801 over frequency;
- 5 Fig. 8 shows a graphical representation of a power spectrum of an audio signal output by block 802 over frequency;
- Fig. 9 shows a block-schematic diagram of an amplitude calculation;
- 10 Fig. 10a shows a block schematic diagram of a modulator;
- Fig. 10b shows a graphical representation of the location of coefficients on the time-frequency claim;
- 15 Figs. 11a and 11b show a block-schematic diagrams of implementation alternatives of the synchronization module;
- Fig. 12a shows a graphical representation of the problem of finding the temporal alignment of a watermark;
- 20 Fig. 12b shows a graphical representation of the problem of identifying the message start;
- Fig. 12c shows a graphical representation of a temporal alignment of synchronization sequences in a full message synchronization mode;
- 25 Fig. 12d shows a graphical representation of the temporal alignment of the synchronization sequences in a partial message synchronization mode;
- 30 Fig. 12e shows a graphical representation of input data of the synchronization module;
- Fig. 12f shows a graphical representation of a concept of identifying a synchronization hit;
- 35 Fig. 12g shows a block-schematic diagram of a synchronization signature correlator;
- Fig. 13a shows a graphical representation of an example for a temporal despreading;

- Fig. 13b shows a graphical representation of an example for an element-wise multiplication between bits and spreading sequences;
- 5 Fig. 13c shows a graphical representation of an output of the synchronization signature correlator after temporal averaging;
- Fig. 13d shows a graphical representation of an output of the synchronization signature correlator filtered with the auto-correlation function of the synchronization signature;
- 10 Fig. 14 shows a block-schematic diagram of a watermark extractor, according to an embodiment of the invention;
- 15 Fig. 15 shows a schematic representation of a selection of a part of the time-frequency-domain representation as a candidate message;
- Fig. 16 shows a block-schematic diagram of an analysis module;
- 20 Fig. 17a shows a graphical representation of an output of a synchronization correlator;
- Fig. 17b shows a graphical representation of decoded messages;
- 25 Fig. 17c shows a graphical representation of a synchronization position, which is extracted from a watermarked signal;
- Fig. 18a shows a graphical representation of a payload, a payload with a Viterbi termination sequence, a Viterbi-encoded payload and a repetition-coded version of the Viterbi-coded payload;
- 30 Fig. 18b shows a graphical representation of subcarriers used for embedding a watermarked signal;
- 35 Fig. 19 shows a graphical representation of an uncoded message, a coded message, a synchronization message and a watermark signal, in which the synchronization sequence is applied to the messages;

- Fig. 20 shows a schematic representation of a first step of a so-called “ABC synchronization” concept;
- 5 Fig. 21 shows a graphical representation of a second step of the so-called “ABC synchronization” concept;
- Fig. 22 shows a graphical representation of a third step of the so-called “ABC synchronization” concept;
- 10 Fig. 23 shows a graphical representation of a message comprising a payload and a CRC portion;
- Fig. 24 shows a block schematic diagram of a watermark signal provider, according to an embodiment of the invention; and
- 15 Fig. 25 shows a flowchart of a method for providing a watermark signal in dependence on a time-frequency domain representation, according to an embodiment of the invention.

20

Detailed Description of the Embodiments

1. Watermark signal provider

25 In the following, a watermark signal provider 2400 will be described taking reference to Fig. 24, which shows a block schematic diagram of such a watermark signal provider.

The watermark signal provider 2400 is configured to receive watermark data, as a time domain frequency representation 2410 at an input and to provide, on the basis thereof, a
30 watermark signal 2420 at an output. The watermark generator 2400 comprises a time-frequency domain waveform provider 2430 and a time domain waveform combiner 2460. The time-frequency domain waveform provider 2430 is configured to provide time domain waveforms 2440 for a plurality of frequency subbands, based on the time-frequency domain representation 2420 of the watermark data. The time-frequency domain waveform
35 provider 2430 is configured to map a given value of the time-frequency domain representation 2410 onto a bit shaping function 2450. A temporal extension of the bit shaping function 2450 is longer than the bit interval associated to the given value of the time-frequency domain representation 2410, such that there is a temporal overlap between

bit shaped functions provided for temporally subsequent values of the time-frequency domain representation 2410 of the same frequency subband. The time-frequency domain waveform provider 2430 is further configured such that a time domain waveform 2440 of a given frequency subband contains a plurality of bit shaped functions provided for temporally subsequent values of the time-frequency domain representation 2410 of the same frequency subband. The time-domain waveform combiner 2460 is configured to combine the provided waveforms 2440 for the plurality of frequencies of the time-frequency domain waveform provider 2430 to derive the watermark signal 2420.

According to an embodiment, the time-frequency domain waveform provider 2430 may comprise a plurality of bit shaping blocks configured to map a given value of the time-frequency domain representation 2410 of the watermark data onto a bit shaping function 2450, the outputs of the bit shaping blocks are therefore bit shaped functions or waveforms in time domain. The time-frequency domain waveform provider 2430 may comprise as many bit shaping blocks as frequency subbands in the time-frequency domain representation of the watermark data.

According to a further embodiment the, watermark signal provider 2400 may comprise a weight tuner. The weight tuner may also be called psychoacoustical processing module. The weight tuner may be configured to tune the weight or an amplitude of bit shaped functions corresponding to values of the time-frequency domain representation 2410 of the watermark data. A weight of a bit shaped function may be tuned such that, as much energy as possible is assigned to a bit shaped function but the watermark signal 2420 is still kept inaudible. The weight tuner may tune the weight in an iterative process for every bit shaped function corresponding to a value of the time-frequency domain representation 2410. Therefore the weights of different bit shaped function can vary.

2. Method for providing a Watermark signal

Fig. 25 shows a method 2500 of providing a watermark signal in dependence on a time-frequency domain representation of watermark data. The method 2500 comprises a first step 2510 of providing time domain waveforms for a plurality of frequency subbands, based on a time-frequency domain representation of watermark data by mapping a given value of the time-frequency domain representation onto a bit shaping function, wherein a temporal extension of the bit shaping function is longer than the bit interval associated to the given value of the time-frequency domain representation, such that there is a temporal overlap between bit shaped functions provided for temporally subsequent values of the time-frequency domain representation of the same frequency subband. A time domain

waveform of a given frequency subband contains a plurality of bit shaped functions provided for temporally subsequent values of the time frequency domain representation of the same frequency subband.

5 The method 2500 further comprises a step 2520 of combining the provided waveforms for the plurality of frequencies to derive the watermark signal. The watermark signal may for example be a sum of the provided waveforms for the plurality of frequencies.

Optionally, the method 2500 may comprise further steps corresponding to the features of the apparatus described above.

10

3. System Description

In the following, a system for a watermark transmission will be described, which comprises a watermark inserter and a watermark decoder. Naturally, the watermark
15 inserter and the watermark decoder can be used independent from each other.

For the description of the system a top-down approach is chosen here. First, it is distinguished between encoder and decoder. Then, in sections 3.1 to 3.5 each processing
20 block is described in detail.

20

The basic structure of the system can be seen in Figures 1 and 2, which depict the encoder and decoder side, respectively. Fig 1 shows a block schematic diagram of a watermark inserter 100. At the encoder side, the watermark signal 101b is generated in the processing
25 block 101 (also designated as watermark generator) from binary data 101a and on the basis of information 104, 105 exchanged with the psychoacoustical processing module 102. The information provided from block 102 typically guarantees that the watermark is inaudible. The watermark generated by the watermark generator 101 is then added to the audio signal 106. The watermarked signal 107 can then be transmitted, stored, or further processed. In
30 case of a multimedia file, e.g., an audio-video file, a proper delay needs to be added to the video stream not to lose audio-video synchronicity. In case of a multichannel audio signal, each channel is processed separately as explained in this document. The processing blocks 101 (watermark generator) and 102 (psychoacoustical processing module) are explained in detail in Sections 3.1 and 3.2, respectively.

35

The decoder side is depicted in Figure 2, which shows a block schematic diagram of a watermark detector 200. A watermarked audio signal 200a, e.g., recorded by a microphone, is made available to the system 200. A first block 203, which is also

designated as an analysis module, demodulates and transforms the data (e.g., the watermarked audio signal) in time/frequency domain (thereby obtaining a time-frequency-domain representation 204 of the watermarked audio signal 200a) passing it to the synchronization module 201, which analyzes the input signal 204 and carries out a temporal synchronization, namely, determines the temporal alignment of the encoded data (e.g. of the encoded watermark data relative to the time-frequency-domain representation). This information (e.g., the resulting synchronization information 205) is given to the watermark extractor 202, which decodes the data (and consequently provides the binary data 202a, which represent the data content of the watermarked audio signal 200a).

10

3.1 The Watermark Generator 101

The watermark generator 101 is depicted detail in Figure 3. Binary data (expressed as ± 1) to be hidden in the audio signal 106 is given to the watermark generator 101. The block 301 organizes the data 101a in packets of equal length M_p . Overhead bits are added (e.g. appended) for signaling purposes to each packet. Let M_s denote their number. Their use will be explained in detail in Section 3.5. Note that in the following each packet of payload bits together with the signaling overhead bits is denoted message.

20

Each message 301a, of length $N_m = M_s + M_p$, is handed over to the processing block 302, the channel encoder, which is responsible of coding the bits for protection against errors. A possible embodiment of this module consists of a convolutional encoder together with an interleaver. The ratio of the convolutional encoder influences greatly the overall degree of protection against errors of the watermarking system. The interleaver, on the other hand, brings protection against noise bursts. The range of operation of the interleaver can be limited to one message but it could also be extended to more messages. Let R_c denote the code ratio, e.g., 1/4. The number of coded bits for each message is N_m/R_c . The channel encoder provides, for example, an encoded binary message 302a.

30

The next processing block, 303, carries out a spreading in frequency domain. In order to achieve sufficient signal to noise ratio, the information (e.g. the information of the binary message 302a) is spread and transmitted in N_f carefully chosen subbands. Their exact position in frequency is decided a priori and is known to both the encoder and the decoder. Details on the choice of this important system parameter is given in Section 3.2.2. The spreading in frequency is determined by the spreading sequence c_f of size $N_f \times 1$. The output 303a of the block 303 consists of N_f bit streams, one for each subband. The i -th bit stream is obtained by multiplying the input bit with the i -th component of spreading

35

sequence c_f . The simplest spreading consists of copying the bit stream to each output stream, namely use a spreading sequence of all ones.

Block 304, which is also designated as a synchronization scheme inserter, adds a synchronization signal to the bit stream. A robust synchronization is important as the decoder does not know the temporal alignment of neither bits nor the data structure, i.e., when each message starts. The synchronization signal consists of N_s sequences of N_f bits each. The sequences are multiplied element wise and periodically to the bit stream (or bit streams 303a). For instance, let \mathbf{a} , \mathbf{b} , and \mathbf{c} , be the $N_s = 3$ synchronization sequences (also designated as synchronization spreading sequences). Block 304 multiplies \mathbf{a} to the first spread bit, \mathbf{b} to the second spread bit, and \mathbf{c} to the third spread bit. For the following bits the process is periodically iterated, namely, \mathbf{a} to the fourth bit, \mathbf{b} for the fifth bit and so on. Accordingly, a combined information-synchronization information 304a is obtained. The synchronization sequences (also designated as synchronization spread sequences) are carefully chosen to minimize the risk of a false synchronization. More details are given in Section 3.4. Also, it should be noted that a sequence \mathbf{a} , \mathbf{b} , \mathbf{c}, \dots may be considered as a sequence of synchronization spread sequences.

Block 305 carries out a spreading in time domain. Each spread bit at the input, namely a vector of length N_f , is repeated in time domain N_t times. Similarly to the spreading in frequency, we define a spreading sequence c_t of size $N_t \times 1$. The i -th temporal repetition is multiplied with the i -th component of c_t .

The operations of blocks 302 to 305 can be put in mathematical terms as follows. Let \mathbf{m} of size $1 \times N_m = R_c$ be a coded message, output of 302. The output 303a (which may be considered as a spread information representation \mathbf{R}) of block 303 is

$$c_f \cdot \mathbf{m} \quad \text{of size} \quad N_f \times N_m / R_c \quad (1)$$

the output 304a of block 304, which may be considered as a combined information-synchronization representation \mathbf{C} , is

$$S \circ (c_f \cdot \mathbf{m}) \quad \text{of size} \quad N_f \times N_m / R_c \quad (2)$$

where \circ denotes the Schur element-wise product and

$$S = [\dots \ a \ b \ c \ \dots \ a \ b \ \dots \] \text{ of size } N_f \times N_m/R_c. \quad (3)$$

The output 305a of 305 is

$$5 \quad (S \circ (c_f \cdot m)) \diamond c_t^T \text{ of size } N_f \times N_t \cdot N_m/R_c \quad (4)$$

where \diamond and T denote the Kronecker product and transpose, respectively. Please recall that binary data is expressed as ± 1 .

10

Block **306** performs a differential encoding of the bits. This step gives the system additional robustness against phase shifts due to movement or local oscillator mismatches. More details on this matter are given in Section 3.3. If $b(i; j)$ is the bit for the i -th frequency band and j -th time block at the input of block 306, the output bit $b_{\text{diff}}(i; j)$ is

$$b_{\text{diff}}(i, j) = b_{\text{diff}}(i, j - 1) \cdot b(i, j). \quad (5)$$

20 At the beginning of the stream, that is for $j = 0$, $b_{\text{diff}}(i, j - 1)$ is set to 1.

Block 307 carries out the actual modulation, i.e., the generation of the watermark signal waveform depending on the binary information 306a given at its input. A more detailed schematics is given in Figure 4. N_f parallel inputs, 401 to $40N_f$ contain the bit streams for the different subbands. Each bit of each subband stream is processed by a bit shaping block (411 to $41N_f$). The output of the bit shaping blocks are waveforms in time domain. The waveform generated for the j -th time block and i -th subband, denoted by $s_{i,j}(t)$, on the basis of the input bit $b_{\text{diff}}(i, j)$

30 is computed as follows

$$s_{i,j}(t) = b_{\text{diff}}(i, j) \gamma(i, j) \cdot g_i(t - j \cdot T_b), \quad (6)$$

35 where $\gamma(i; j)$ is a weighting factor provided by the psychoacoustical processing unit 102, T_b is the bit time interval, and $g_i(t)$ is the bit forming function for the i -th subband. The bit

forming function is obtained from a baseband function $g_i^T(t)$ modulated in frequency with a cosine

$$g_i(t) = g_i^T(t) \cdot \cos(2\pi f_i t) \quad (7)$$

5

where f_i is the center frequency of the i -th subband and the superscript T stands for transmitter. The baseband functions can be different for each subband. If chosen identical, a more efficient implementation at the decoder is possible. See Section 3.3 for more details.

10

The bit shaping for each bit is repeated in an iterative process controlled by the psychoacoustical processing module (102). Iterations are necessary to fine tune the weights $\gamma(i, j)$ to assign as much energy as possible to the watermark while keeping it inaudible. More details are given in Section 3.2.

15

The complete waveform at the output of the i -th bit shaping filter is

$$s_i(t) = \sum_j s_{i,j}(t). \quad (8)$$

20

The bit forming baseband function $g_i^T(t)$ is normally non zero for a time interval much larger than T_b , although the main energy is concentrated within the bit interval. An example can be seen in Figure 12a where the same bit forming baseband function is plotted for two adjacent bits. In the figure we have $T_b = 40$ ms. The choice of T_b as well as the shape of the function affect the system considerably. In fact, longer symbols provide narrower frequency responses. This is particularly beneficial in reverberant environments. In fact, in such scenarios the watermarked signal reaches the microphone via several propagation paths, each characterized by a different propagation time. The resulting channel exhibits strong frequency selectivity. Interpreted in time domain, longer symbols are beneficial as echoes with a delay comparable to the bit interval yield constructive interference, meaning that they increase the received signal energy. Notwithstanding, longer symbols bring also a few drawbacks; larger overlaps might lead to intersymbol interference (ISI) and are for sure more difficult to hide in the audio signal, so that the psychoacoustical processing module would allow less energy than for shorter symbols.

30

The watermark signal is obtained by summing all outputs of the bit shaping filters

35

$$\sum_i s_i(t).$$

(9)

5 3.2 The Psychoacoustical Processing Module 102

As depicted in Figure 5, the psychoacoustical processing module 102 consists of 3 parts. The first step is an analysis module **501** which transforms the time audio signal into the time/frequency domain. This analysis module may carry out parallel analyses in different
 10 time/frequency resolutions. After the analysis module, the time/frequency data is transferred to the psychoacoustic model (PAM) **502**, in which masking thresholds for the watermark signal are calculated according to psychoacoustical considerations (see E. Zwicker H.Fastl, "Psychoacoustics Facts and models"). The masking thresholds indicate the amount of energy which can be hidden in the audio signal for each subband and time
 15 block. The last block in the psychoacoustical processing module 102 depicts the amplitude calculation module **503**. This module determines the amplitude gains to be used in the generation of the watermark signal so that the masking thresholds are satisfied, i.e., the embedded energy is less or equal to the energy defined by the masking thresholds.

20

3.2.1 The Time/Frequency Analysis 501

Block **501** carries out the time/frequency transformation of the audio signal by means of a lapped transform. The best audio quality can be achieved when multiple time/frequency
 25 resolutions are performed. One efficient embodiment of a lapped transform is the short time Fourier transform (STFT), which is based on fast Fourier transforms (FFT) of windowed time blocks. The length of the window determines the time/frequency resolution, so that longer windows yield lower time and higher frequency resolutions, while shorter windows vice versa. The shape of the window, on the other hand, among
 30 other things, determines the frequency leakage.

For the proposed system, we achieve an inaudible watermark by analyzing the data with two different resolutions. A first filter bank is characterized by a hop size of T_b , i.e., the bit length. The hop size is the time interval between two adjacent time blocks. The window
 35 length is approximately T_b . Please note that the window shape does not have to be the same as the one used for the bit shaping, and in general should model the human hearing system. Numerous publications study this problem.

The second filter bank applies a shorter window. The higher temporal resolution achieved is particularly important when embedding a watermark in speech, as its temporal structure is in general finer than T_b .

5

The sampling rate of the input audio signal is not important, as long as it is large enough to describe the watermark signal without aliasing. For instance, if the largest frequency component contained in the watermark signal is 6 kHz, then the sampling rate of the time signals must be at least 12 kHz.

10

3.2.2 The Psychoacoustical Model 502

The psychoacoustical model 502 has the task to determine the masking thresholds, i.e., the amount of energy which can be hidden in the audio signal for each subband and time block keeping the watermarked audio signal indistinguishable from the original.

15

The i -th subband is defined between two limits, namely $f_i^{(\min)}$ and $f_i^{(\max)}$. The subbands are determined by defining N_f center frequencies f_i and letting $f_{i-1}^{(\max)} = f_i^{(\min)}$ for $i = 2, 3, \dots, N_f$. An appropriate choice for the center frequencies is given by the Bark scale proposed by Zwicker in 1961. The subbands become larger for higher center frequencies. A possible implementation of the system uses 9 subbands ranging from 1.5 to 6 kHz arranged in an appropriate way.

20

The following processing steps are carried out separately for each time/frequency resolution for each subband and each time block. The processing step **801** carries out a spectral smoothing. In fact, tonal elements, as well as notches in the power spectrum need to be smoothed. This can be carried out in several ways. A tonality measure may be computed and then used to drive an adaptive smoothing filter. Alternatively, in a simpler implementation of this block, a median-like filter can be used. The median filter considers a vector of values and outputs their median value. In a median-like filter the value corresponding to a different quantile than 50% can be chosen. The filter width is defined in Hz and is applied as a non-linear moving average which starts at the lower frequencies and ends up at the highest possible frequency. The operation of **801** is illustrated in Figure 7. The red curve is the output of the smoothing.

25

30

35

Once the smoothing has been carried out, the thresholds are computed by block **802** considering only frequency masking. Also in this case there are different possibilities. One

way is to use the minimum for each subband to compute the masking energy E_i . This is the equivalent energy of the signal which effectively operates a masking. From this value we can simply multiply a certain scaling factor to obtain the masked energy J_i . These factors are different for each subband and time/frequency resolution and are obtained via empirical psychoacoustical experiments. These steps are illustrated in Figure 8.

In block **805**, temporal masking is considered. In this case, different time blocks for the same subband are analyzed. The masked energies J_i are modified according to an empirically derived postmasking profile. Let us consider two adjacent time blocks, namely $k-1$ and k . The corresponding masked energies are $J_i(k-1)$ and $J_i(k)$. The postmasking profile defines that, e.g., the masking energy E_i can mask an energy J_i at time k and $\alpha \cdot J_i$ at time $k+1$. In this case, block **805** compares $J_i(k)$ (the energy masked by the current time block) and $\alpha \cdot J_i(k+1)$ (the energy masked by the previous time block) and chooses the maximum. Postmasking profiles are available in the literature and have been obtained via empirical psychoacoustical experiments. Note that for large T_b , i.e., > 20 ms, postmasking is applied only to the time/frequency resolution with shorter time windows.

Summarizing, at the output of block **805** we have the masking thresholds per each subband and time block obtained for two different time/frequency resolutions. The thresholds have been obtained by considering both frequency and time masking phenomena. In block **806**, the thresholds for the different time/frequency resolutions are merged. For instance, a possible implementation is that **806** considers all thresholds corresponding to the time and frequency intervals in which a bit is allocated, and chooses the minimum.

3.2.3 The Amplitude Calculation Block 503

Please refer to Figure 9. The input of 503 are the thresholds 505 from the psychoacoustical model 502 where all psychoacoustics motivated calculations are carried out. In the amplitude calculator 503 additional computations with the thresholds are performed. First, an amplitude mapping **901** takes place. This block merely converts the masking thresholds (normally expressed as energies) into amplitudes which can be used to scale the bit shaping function defined in Section 3.1. Afterwards, the amplitude adaptation block **902** is run. This block iteratively adapts the amplitudes $\gamma(i, j)$ which are used to multiply the bit shaping functions in the watermark generator 101 so that the masking thresholds are indeed fulfilled. In fact, as already discussed, the bit shaping function normally extends for a time interval larger than T_b . Therefore, multiplying the correct amplitude $\gamma(i, j)$ which fulfills the masking threshold at point i, j does not necessarily fulfill the requirements at point $i, j-1$. This is particularly crucial at strong onsets, as a preecho becomes audible.

Another situation which needs to be avoided is the unfortunate superposition of the tails of different bits which might lead to an audible watermark. Therefore, block 902 analyzes the signal generated by the watermark generator to check whether the thresholds have been fulfilled. If not, it modifies the amplitudes $\gamma(i, j)$ accordingly.

5

This concludes the encoder side. The following sections deal with the processing steps carried out at the receiver (also designated as watermark decoder).

10 3.3 The Analysis Module 203

The analysis module 203 is the first step (or block) of the watermark extraction process. Its purpose is to transform the watermarked audio signal 200a back into N_f bit streams $\hat{b}_i(j)$ (also designated with 204), one for each spectral subband i . These are further processed by the synchronization module 201 and the watermark extractor 202, as discussed in Sections 15 3.4 and 3.5, respectively. Note that the $\hat{b}_i(j)$ are soft bit streams, i.e., they can take, for example, any real value and no hard decision on the bit is made yet.

The analysis module consists of three parts which are depicted in Figure 16: The analysis 20 filter bank 1600, the amplitude normalization block 1604 and the differential decoding 1608.

3.3.1 Analysis filter bank 1600

25

The watermarked audio signal is transformed into the time-frequency domain by the analysis filter bank 1600 which is shown in detail in Figure 10a. The input of the filter bank is the received watermarked audio signal $r(t)$. Its output are the complex coefficients $b_i^{AFB}(j)$ for the i -th branch or subband at time instant j . These values contain information 30 about the amplitude and the phase of the signal at center frequency f_i and time $j \cdot T_b$.

The filter bank 1600 consists of N_f branches, one for each spectral subband i . Each branch splits up into an upper subbranch for the in-phase component and a lower subbranch for the quadrature component of the subband i . Although the modulation at the watermark generator and thus the watermarked audio signal are purely real-valued, the complex-valued analysis of the signal at the receiver is needed because rotations of the modulation constellation introduced by the channel and by synchronization misalignments are not 35

known at the receiver. In the following we consider the i -th branch of the filter bank. By combining the in-phase and the quadrature subbranch, we can define the complex-valued baseband signal $b_i^{\text{AFB}}(t)$ as

$$b_i^{\text{AFB}}(t) = r(t) \cdot e^{-j2\pi f_i t} * g_i^{\text{R}}(t) \quad (10)$$

where $*$ indicates convolution and $g_i^{\text{R}}(t)$ is the impulse response of the receiver lowpass filter of subband i . Usually $g_i^{\text{R}}(t)$ is equal to the baseband bit forming function $g_i^{\text{T}}(t)$ of subband i in the modulator 307 in order to fulfill the matched filter condition, but other impulse responses are possible as well.

In order to obtain the coefficients $b_i^{\text{AFB}}(j)$ with rate $1/T_b$, the continuous output $b_i^{\text{AFB}}(t)$ must be sampled. If the correct timing of the bits was known by the receiver, sampling with rate $1/T_b$ would be sufficient. However, as the bit synchronization is not known yet, sampling is carried out with rate N_{os}/T_b where N_{os} is the analysis filter bank oversampling factor. By choosing N_{os} sufficiently large (e.g. $N_{\text{os}} = 4$), we can assure that at least one sampling cycle is close enough to the ideal bit synchronization. The decision on the best oversampling layer is made during the synchronization process, so all the oversampled data is kept until then. This process is described in detail in Section 3.4.

At the output of the i -th branch we have the coefficients $b_i^{\text{AFB}}(j, k)$, where j indicates the bit number or time instant and k indicates the oversampling position within this single bit, where $k = 1; 2; \dots, N_{\text{os}}$.

Figure 10b gives an exemplary overview of the location of the coefficients on the time-frequency plane. The oversampling factor is $N_{\text{os}} = 2$. The height and the width of the rectangles indicate respectively the bandwidth and the time interval of the part of the signal that is represented by the corresponding coefficient $b_i^{\text{AFB}}(j, k)$.

If the subband frequencies f_i are chosen as multiples of a certain interval Δf the analysis filter bank can be efficiently implemented using the Fast Fourier Transform (FFT).

Without loss of generality and to simplify the description, we assume that the bit synchronization is known and that $N_{os} = 1$ in the following. That is, we have complex coefficients $b_i^{AFB}(j)$ at the input of the normalization block 1604. As no channel state information is available at the receiver (i.e., the propagation channel is unknown), an equal gain combining (EGC) scheme is used. Due to the time and frequency dispersive channel, the energy of the sent bit $b_i(j)$ is not only found around the center frequency f_i and time instant j , but also at adjacent frequencies and time instants. Therefore, for a more precise weighting, additional coefficients at frequencies $f_i \pm n \Delta f$ are calculated and used for normalization of coefficient $b_i^{AFB}(j)$. If $n = 1$ we have, for example,

$$b_i^{norm}(j) = \frac{b_i^{AFB}(j)}{\sqrt{1/3 \cdot (|b_i^{AFB}(j)|^2 + |b_{i-\Delta f}^{AFB}(j)|^2 + |b_{i+\Delta f}^{AFB}(j)|^2)}} \quad (11)$$

The normalization for $n > 1$ is a straightforward extension of the formula above. In the same fashion we can also choose to normalize the soft bits by considering more than one time instant. The normalization is carried out for each subband i and each time instant j . The actual combining of the EGC is done at later steps of the extraction process.

3.3.3 Differential decoding 1608

At the input of the differential decoding block 1608 we have amplitude normalized complex coefficients $b_i^{norm}(j)$ which contain information about the phase of the signal components at frequency f_i and time instant j . As the bits are differentially encoded at the transmitter, the inverse operation must be performed here. The soft bits $\widehat{b}_i(j)$ are obtained by first calculating the difference in phase of two consecutive coefficients and then taking the real part:

$$\widehat{b}_i(j) = \text{Re}\{b_i^{norm}(j) \cdot b_i^{norm*}(j-1)\} \quad (12)$$

$$= \text{Re}\{|b_i^{norm}(j)| \cdot |b_i^{norm}(j-1)| \cdot e^{j(\varphi_j - \varphi_{j-1})}\} \quad (13)$$

This has to be carried out separately for each subband because the channel normally introduces different phase rotations in each subband.

3.4 The Synchronization Module 201

5 The synchronization module's task is to find the temporal alignment of the watermark. The problem of synchronizing the decoder to the encoded data is twofold. In a first step, the analysis filterbank must be aligned with the encoded data, namely the bit shaping functions $g_i^T(t)$ used in the synthesis in the modulator must be aligned with the filters $g_i^R(t)$ used for the analysis. This problem is illustrated in Figure 12a, where the analysis filters are
 10 identical to the synthesis ones. At the top, three bits are visible. For simplicity, the waveforms for all three bits are not scaled. The temporal offset between different bits is T_b . The bottom part illustrates the synchronization issue at the decoder: the filter can be applied at different time instants, however, only the position marked in red (curve 1299a) is correct and allows to extract the first bit with the best signal to noise ratio SNR and
 15 signal to interference ratio SIR. In fact, an incorrect alignment would lead to a degradation of both SNR and SIR. We refer to this first alignment issue as "bit synchronization". Once the bit synchronization has been achieved, bits can be extracted optimally. However, to correctly decode a message, it is necessary to know at which bit a new message starts. This issue is illustrated in Figure 12b and is referred to as message synchronization. In the
 20 stream of decoded bits only the starting position marked in red (position 1299b) is correct and allows to decode the k-th message.

We first address the message synchronization only. The synchronization signature, as explained in Section 3.1, is composed of N_s sequences in a predetermined order which are
 25 embedded continuously and periodically in the watermark. The synchronization module is capable of retrieving the temporal alignment of the synchronization sequences. Depending on the size N_s we can distinguish between two modes of operation, which are depicted in Figure 12c and 12d, respectively.

30 In the full message synchronization mode (Fig. 12c) we have $N_s = N_m/R_c$. For simplicity in the figure we assume $N_s = N_m/R_c = 6$ and no time spreading, i.e., $N_t = 1$. The synchronization signature used, for illustration purposes, is shown beneath the messages. In reality, they are modulated depending on the coded bits and frequency spreading sequences, as explained in Section 3.1. In this mode, the periodicity of the synchronization
 35 signature is identical to the one of the messages. The synchronization module therefore can identify the beginning of each message by finding the temporal alignment of the synchronization signature. We refer to the temporal positions at which a new

synchronization signature starts as synchronization hits. The synchronization hits are then passed to the watermark extractor 202.

The second possible mode, the partial message synchronization mode (Fig. 12d), is depicted in Figure 12d. In this case we have $N_s < N_m = R_c$. In the figure we have taken $N_s = 3$, so that the three synchronization sequences are repeated twice for each message. Please note that the periodicity of the messages does not have to be multiple of the periodicity of the synchronization signature. In this mode of operation, not all synchronization hits correspond to the beginning of a message. The synchronization module has no means of distinguishing between hits and this task is given to the watermark extractor 202.

The processing blocks of the synchronization module are depicted in Figures 11a and 11b. The synchronization module carries out the bit synchronization and the message synchronization (either full or partial) at once by analyzing the output of the synchronization signature correlator 1201. The data in time/frequency domain 204 is provided by the analysis module. As the bit synchronization is not yet available, block 203 oversamples the data with factor N_{os} , as described in Section 3.3. An illustration of the input data is given in Figure 12e. For this example we have taken $N_{os} = 4$, $N_t = 2$, and $N_s = 3$. In other words, the synchronization signature consists of 3 sequences (denoted with a, b, and c). The time spreading, in this case with spreading sequence $c_t = [1 \ 1]^T$, simply repeats each bit twice in time domain. The exact synchronization hits are denoted with arrows and correspond to the beginning of each synchronization signature. The period of the synchronization signature is $N_t \cdot N_{os} \cdot N_s = N_{sbl}$ which is $2 \cdot 4 \cdot 3 = 24$, for example. Due to the periodicity of the synchronization signature, the synchronization signature correlator (1201) arbitrarily divides the time axis in blocks, called search blocks, of size N_{sbl} , whose subscript stands for search block length. Every search block must contain (or typically contains) one synchronization hit as depicted in Figure 12f. Each of the N_{sbl} bits is a candidate synchronization hit. Block 1201's task is to compute a likelihood measure for each of candidate bit of each block. This information is then passed to block 1204 which computes the synchronization hits.

3.4.1 The synchronization signature correlator 1201

For each of the N_{sbl} candidate synchronization positions the synchronization signature correlator computes a likelihood measure, the latter is larger the more probable it is that the temporal alignment (both bit and partial or full message synchronization) has been found. The processing steps are depicted in Figure 12g.

Accordingly, a sequence 1201a of likelihood values, associated with different positional choices, may be obtained.

5 Block 1301 carries out the temporal despreading, i.e., multiplies every N_t bits with the temporal spreading sequence c_t and then sums them. This is carried out for each of the N_f frequency subbands. Figure 13a shows an example. We take the same parameters as described in the previous section, namely $N_{os} = 4$, $N_t = 2$, and $N_s = 3$. The candidate synchronization position is marked. From that bit, with N_{os} offset, $N_t \cdot N_s$ are taken by
10 block 1301 and time despread with sequence c_t , so that N_s bits are left.

In block 1302 the bits are multiplied element-wise with the N_s spreading sequences (see Figure 13b).

15 In block 1303 the frequency despreading is carried out, namely, each bit is multiplied with the spreading sequence c_f and then summed along frequency.

At this point, if the synchronization position were correct, we would have N_s decoded bits. As the bits are not known to the receiver, block 1304 computes the likelihood measure by
20 taking the absolute values of the N_s values and sums.

The output of block 1304 is in principle a non coherent correlator which looks for the synchronization signature. In fact, when choosing a small N_s , namely the partial message synchronization mode, it is possible to use synchronization sequences (e.g. **a**, **b**, **c**) which
25 are mutually orthogonal. In doing so, when the correlator is not correctly aligned with the signature, its output will be very small, ideally zero. When using the full message synchronization mode it is advised to use as many orthogonal synchronization sequences as possible, and then create a signature by carefully choosing the order in which they are used. In this case, the same theory can be applied as when looking for spreading sequences
30 with good auto correlation functions. When the correlator is only slightly misaligned, then the output of the correlator will not be zero even in the ideal case, but anyway will be smaller compared to the perfect alignment, as the analysis filters cannot capture the signal energy optimally.

35

3.4.2 Synchronization hits computation 1204

This block analyzes the output of the synchronization signature correlator to decide where the synchronization positions are. Since the system is fairly robust against misalignments of up to $T_b/4$ and the T_b is normally taken around 40 ms, it is possible to integrate the output of 1201 over time to achieve a more stable synchronization. A possible implementation of this is given by an IIR filter applied along time with a exponentially decaying impulse response. Alternatively, a traditional FIR moving average filter can be applied. Once the averaging has been carried out, a second correlation along different $N_t \cdot N_s$ is carried out (“different positional choice”). In fact, we want to exploit the information that the autocorrelation function of the synchronization function is known. This corresponds to a Maximum Likelihood estimator. The idea is shown in Figure 13c. The curve shows the output of block 1201 after temporal integration. One possibility to determine the synchronization hit is simply to find the maximum of this function. In Figure 13d we see the same function (in black) filtered with the autocorrelation function of the synchronization signature. The resulting function is plotted in red. In this case the maximum is more pronounced and gives us the position of the synchronization hit. The two methods are fairly similar for high SNR but the second method performs much better in lower SNR regimes. Once the synchronization hits have been found, they are passed to the watermark extractor 202 which decodes the data.

In some embodiments, in order to obtain a robust synchronization signal, synchronization is performed in partial message synchronization mode with short synchronization signatures. For this reason many decodings have to be done, increasing the risk of false positive message detections. To prevent this, in some embodiments signaling sequences may be inserted into the messages with a lower bit rate as a consequence.

This approach is a solution to the problem arising from a sync signature shorter than the message, which is already addressed in the above discussion of the enhanced synchronization. In this case, the decoder doesn't know where a new message starts and attempts to decode at several synchronization points. To distinguish between legitimate messages and false positives, in some embodiments a signaling word is used (i.e. payload is sacrificed to embed a known control sequence). In some embodiments, a plausibility check is used (alternatively or in addition) to distinguish between legitimate messages and false positives.

3.5 The Watermark Extractor 202

The parts constituting the watermark extractor 202 are depicted in Figure 14. This has two inputs, namely 204 and 205 from blocks 203 and 201, respectively. The synchronization

module 201 (see Section 3.4) provides synchronization timestamps, i.e., the positions in time domain at which a candidate message starts. More details on this matter are given in Section 3.4. The analysis filterbank block 203, on the other hand, provides the data in time/frequency domain ready to be decoded.

5

The first processing step, the data selection block 1501, selects from the input 204 the part identified as a candidate message to be decoded. Figure 15b shows this procedure graphically. The input 204 consists of N_f streams of real values. Since the time alignment is not known to the decoder a priori, the analysis block 203 carries out a frequency analysis with a rate higher than $1/T_b$ Hz (oversampling). In Figure 15b we have used an
10 oversampling factor of 4, namely, 4 vectors of size $N_f \times 1$ are output every T_b seconds. When the synchronization block 201 identifies a candidate message, it delivers a timestamp 205 indicating the starting point of a candidate message. The selection block 1501 selects the information required for the decoding, namely a matrix of size $N_f \times N_m/R_c$.
15 This matrix 1501a is given to block 1502 for further processing.

Blocks 1502, 1503, and 1504 carry out the same operations of blocks 1301, 1302, and 1303 explained in Section 3.4.

20 An alternative embodiment of the invention consists in avoiding the computations done in 1502-1504 by letting the synchronization module deliver also the data to be decoded. Conceptually it is a detail. From the implementation point of view, it is just a matter of how the buffers are realized. In general, redoing the computations allows us to have smaller buffers.

25

The channel decoder 1505 carries out the inverse operation of block 302. If channel encoder, in a possible embodiment of this module, consisted of a convolutional encoder together with an interleaver, then the channel decoder would perform the deinterleaving and the convolutional decoding, e.g., with the well known Viterbi algorithm. At the output
30 of this block we have N_m bits, i.e., a candidate message.

Block 1506, the signaling and plausibility block, decides whether the input candidate message is indeed a message or not. To do so, different strategies are possible.

35 The basic idea is to use a signaling word (like a CRC sequence) to distinguish between true and false messages. This however reduces the number of bits available as payload. Alternatively we can use plausibility checks. If the messages for instance contain a

timestamp, consecutive messages must have consecutive timestamps. If a decoded message possesses a timestamp which is not the correct order, we can discard it.

5 When a message has been correctly detected the system may choose to apply the look ahead and/or look back mechanisms. We assume that both bit and message synchronization have been achieved. Assuming that the user is not zapping, the system "looks back" in time and attempts to decode the past messages (if not decoded already) using the same synchronization point (look back approach). This is particularly useful when the system starts. Moreover, in bad conditions, it might take 2 messages to achieve
10 synchronization. In this case, the first message has no chance. With the look back option we can save "good" messages which have not been received only due to back synchronization. The look ahead is the same but works in the future. If we have a message now we know where the next message should be, and we can attempt to decode it anyhow.

15

3.6. Synchronization Details

For the encoding of a payload, for example, a Viterbi algorithm may be used. Fig. 18a shows a graphical representation of a payload 1810, a Viterbi termination sequence 1820, a
20 Viterbi encoded payload 1830 and a repetition-coded version 1840 of the Viterbi-coded payload. For example, the payload length may be 34 bits and the Viterbi termination sequence may comprise 6 bits. If, for example a Viterbi code rate of 1/7 may be used the Viterbi-coded payload may comprise $(34+6)*7=280$ bits. Further, by using a repetition coding of 1/2, the repetition coded version 1840 of the Viterbi-encoded payload 1830 may
25 comprise $280*2=560$ bits. In this example, considering a bit time interval of 42.66 ms, the message length would be 23.9 s. The signal may be embedded with, for example, 9 subcarriers (e.g. placed according to the critical bands) from 1.5 to 6 kHz as indicated by the frequency spectrum shown in Fig. 18b. Alternatively, also another number of subcarriers (e.g. 4, 6, 12, 15 or a number between 2 and 20) within a frequency range
30 between 0 and 20 kHz maybe used.

Fig. 19 shows a schematic illustration of the basic concept 1900 for the synchronization, also called ABC synch. It shows a schematic illustration of an uncoded messages 1910, a coded message 1920 and a synchronization sequence (synch sequence) 1930 as well as the
35 application of the synch to several messages 1920 following each other.

The synchronization sequence or synch sequence mentioned in connection with the explanation of this synchronization concept (shown in Fig. 19 – 23) may be equal to the synchronization signature mentioned before.

5 Further, Fig. 20 shows a schematic illustration of the synchronization found by correlating with the synch sequence. If the synchronization sequence 1930 is shorter than the message, more than one synchronization point 1940 (or alignment time block) may be found within a single message. In the example shown in Fig. 20, 4 synchronization points are found within each message. Therefore, for each synchronization found, a Viterbi decoder (a
10 Viterbi decoding sequence) may be started. In this way, for each synchronization point 1940 a message 2110 may be obtained, as indicated in Fig. 21.

Based on these messages the true messages 2210 may be identified by means of a CRC sequence (cyclic redundancy check sequence) and/or a plausibility check, as shown in Fig.
15 22.

The CRC detection (cyclic redundancy check detection) may use a known sequence to identify true messages from false positive. Fig. 23 shows an example for a CRC sequence added to the end of a payload.

20 The probability of false positive (a message generated based on a wrong synchronization point) may depend on the length of the CRC sequence and the number of Viterbi decoders (number of synchronization points within a single message) started. To increase the length of the payload without increasing the probability of false positive a plausibility may be exploited (plausibility test) or the length of the synchronization sequence (synchronization
25 signature) may be increased.

4. Concepts and Advantages

30 In the following, some aspects of the above discussed system will be described, which are considered as being innovative. Also, the relation of those aspects to the state-of-the-art technologies will be discussed.

35

4.1. Continuous synchronization

Some embodiments allow for a continuous synchronization. The synchronization signal, which we denote as synchronization signature, is embedded continuously and parallel to the data via multiplication with sequences (also designated as synchronization spread sequences) known to both transmit and receive side.

5

Some conventional systems use special symbols (other than the ones used for the data), while some embodiments according to the invention do not use such special symbols. Other classical methods consist of embedding a known sequence of bits (preamble) time-multiplexed with the data, or embedding a signal frequency-multiplexed with the data.

10

However, it has been found that using dedicated sub-bands for synchronization is undesired, as the channel might have notches at those frequencies, making the synchronization unreliable. Compared to the other methods, in which a preamble or a special symbol is time-multiplexed with the data, the method described herein is more advantageous as the method described herein allows to track changes in the synchronization (due e.g. to movement) continuously.

15

Furthermore, the energy of the watermark signal is unchanged (e.g. by the multiplicative introduction of the watermark into the spread information representation), and the synchronization can be designed independent from the psychoacoustical model and data rate. The length in time of the synchronization signature, which determines the robustness of the synchronization, can be designed at will completely independent of the data rate.

20

Another classical method consists of embedding a synchronization sequence code-multiplexed with the data. When compared to this classical method, the advantage of the method described herein is that the energy of the data does not represent an interfering factor in the computation of the correlation, bringing more robustness. Furthermore, when using code-multiplexing, the number of orthogonal sequences available for the synchronization is reduced as some are necessary for the data.

25

30

To summarize, the continuous synchronization approach described herein brings along a large number of advantages over the conventional concepts.

35

However, in some embodiments according to the invention, a different synchronization concept may be applied.

4.2. 2D spreading

Some embodiments of the proposed system carry out spreading in both time and frequency domain, i.e. a 2-dimensional spreading (briefly designated as 2D-spreading). It has been found that this is advantageous with respect to 1D systems as the bit error rate can be
5 further reduced by adding redundancy in e.g. time domain.

However, in some embodiments according to the invention, a different spreading concept may be applied.

10

4.3. Differential encoding and Differential decoding

In some embodiments according to the invention, an increased robustness against movement and frequency mismatch of the local oscillators (when compared to
15 conventional systems) is brought by the differential modulation. It has been found that in fact, the Doppler effect (movement) and frequency mismatches lead to a rotation of the BPSK constellation (in other words, a rotation on the complex plane of the bits). In some embodiments, the detrimental effects of such a rotation of the BPSK constellation (or any other appropriate modulation constellation) are avoided by using a differential encoding or
20 differential decoding.

However, in some embodiments according to the invention, a different encoding concept or decoding concept may be applied. Also, in some cases, the differential encoding may be
25 omitted.

25

4.4. Bit shaping

In some embodiments according to the invention, bit shaping brings along a significant
30 improvement of the system performance, because the reliability of the detection can be increased using a filter adapted to the bit shaping.

In accordance with some embodiments, the usage of bit shaping with respect to watermarking brings along improved reliability of the watermarking process. It has been
35 found that particularly good results can be obtained if the bit shaping function is longer than the bit interval.

However, in some embodiments according to the invention, a different bit shaping concept may be applied. Also, in some cases, the bit shaping may be omitted.

5 4.5. Interactive between Psychoacoustic Model (PAM) and Filter Bank (FB) synthesis

In some embodiments, the psychoacoustical model interacts with the modulator to fine tune the amplitudes which multiply the bits.

10 However, in some other embodiments, this interaction may be omitted.

4.6. Look ahead and look back features

In some embodiments, so called "Look back" and "look ahead" approaches are applied.

15

In the following, these concepts will be briefly summarized. When a message is correctly decoded, it is assumed that synchronization has been achieved. Assuming that the user is not zapping, in some embodiments a look back in time is performed and it is tried to decode the past messages (if not decoded already) using the same synchronization point (look back approach). This is particularly useful when the system starts.

20

In bad conditions, it might take 2 messages to achieve synchronization. In this case, the first message has no chance in conventional systems. With the look back option, which is used in some embodiments of the invention, it is possible to save (or decode) "good" messages which have not been received only due to back synchronization.

25

The look ahead is the same but works in the future. If I have a message now I know where my next message should be, and I can try to decode it anyhow. Accordingly, overlapping messages can be decoded.

30

However, in some embodiments according to the invention, the look ahead feature and/or the look back feature may be omitted.

35 4.7. Increased synchronization robustness

In some embodiments, in order to obtain a robust synchronization signal, synchronization is performed in partial message synchronization mode with short synchronization

signatures. For this reason many decodings have to be done, increasing the risk of false positive message detections. To prevent this, in some embodiments signaling sequences may be inserted into the messages with a lower bit rate as a consequence.

- 5 However, in some embodiments according to the invention, a different concept for improving the synchronization robustness may be applied. Also, in some cases, the usage of any concepts for increasing the synchronization robustness may be omitted.

10 4.8. Other enhancements

In the following, some other general enhancements of the above described system with respect to background art will be put forward and discussed:

- 15 1. lower computational complexity
2. better audio quality due to the better psychoacoustical model
3. more robustness in reverberant environments due to the narrowband multicarrier
20 signals
4. an SNR estimation is avoided in some embodiments. This allows for better robustness, especially in low SNR regimes.

25 Some embodiments according to the invention are better than conventional systems, which use very narrow bandwidths of, for example, 8Hz for the following reasons:

1. 8 Hz bandwidths (or a similar very narrow bandwidth) requires very long time
symbols because the psychoacoustical model allows very little energy to make it inaudible;
- 30 2. 8 Hz (or a similar very narrow bandwidth) makes it sensitive against time varying Doppler spectra. Accordingly, such a narrow band system is typically not good enough if implemented, e.g., in a watch.

35 Some embodiments according to the invention are better than other technologies for the following reasons:

1. Techniques which input an echo fail completely in reverberant rooms. In contrast, in some embodiments of the invention, the introduction of an echo is avoided.

2. Techniques which use only time spreading have longer message duration in comparison embodiments of the above described system in which a two-dimensional spreading, for example both in time and in frequency, is used.

5

Some embodiments according to the invention are better than the system described in DE 196 40 814, because one or more of the following disadvantages of the system according to said document are overcome:

10

- the complexity in the decoder according to DE 196 40 814 is very high, a filter of length $2N$ with $N = 128$ is used
- the system according to DE 196 40 814 comprises a long message duration
- in the system according to DE 196 40 814 spreading only in time domain with relatively high spreading gain (e.g. 128)
- in the system according to DE 196 40 814 the signal is generated in time domain, transformed to spectral domain, weighted, transformed back to time domain, and superposed to audio, which makes the system very complex

15

20

5. Applications

The invention comprises a method to modify an audio signal in order to hide digital data and a corresponding decoder capable of retrieving this information while the perceived quality of the modified audio signal remains indistinguishable to the one of the original.

25

Examples of possible applications of the invention are given in the following:

1. Broadcast monitoring: a watermark containing information on e.g. the station and time is hidden in the audio signal of radio or television programs. Decoders, incorporated in small devices worn by test subjects, are capable to retrieve the watermark, and thus collect valuable information for advertisements agencies, namely who watched which program and when.

30

2. Auditing: a watermark can be hidden in, e.g., advertisements. By automatically monitoring the transmissions of a certain station it is then possible to know when exactly the ad was broadcast. In a similar fashion it is possible to retrieve statistical information

35

about the programming schedules of different radios, for instance, how often a certain music piece is played, etc.

3. Metadata embedding: the proposed method can be used to hide digital information about the music piece or program, for instance the name and author of the piece or the duration of the program etc.

6. Implementation Alternatives

10

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.

15

- 20 The inventive encoded watermark signal, or an audio signal into which the watermark signal is embedded, can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

- 25 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.

30

- 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.

35

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.

5

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

10

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.

15

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.

20

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.

25

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.

30

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.

35

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. A watermark signal provider (2400; 307) for providing a watermark signal (2420, wms(t); 307a; 101b) in dependence on a time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of watermark data, in which the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) comprises values associated to frequency subbands (i) and bit intervals (j), the watermark signal provider (2400; 307) comprising:
- 10 a time-frequency-domain waveform provider (2430; 411-41N_f, 421-42N_f) configured to provide time-domain waveforms (2440; $s_i(t)$) for a plurality of frequency subbands (i), based on the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the watermark data, wherein the time-frequency-domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured to map a given value ($b_{\text{diff}}(i, j)$) of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) onto a bit shaping function ($g_i(t)$), wherein a temporal extension of the bit shaping function ($g_i(t)$) is longer than the bit interval (j) associated to the given value ($b_{\text{diff}}(i, j)$) of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that there is a temporal overlap between bit shaped functions ($g_i(t)$) provided for temporally subsequent values of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency subband (i); and
- 25 wherein the time-frequency-domain waveform provider (2430; 411-41N_f, 421-42N_f) is further configured such that a time-domain waveform (2440, $s_i(t)$) of a given frequency subband (i) contains a plurality of bit shaped functions ($s_{i,j}(t)$) provided for temporally subsequent values of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency band (i); and
- 30 a time-domain waveform combiner (2460), to combine the provided time-domain waveforms (2440, $s_i(t)$) for the plurality of frequencies (i) of the time-frequency-domain provider (2430; 411-41N_f, 421-42N_f) to derive the watermark signal (2420, wms(t); 307a; 101b).
- 35 2. The watermark signal provider (2400; 307) according to claim 1, wherein the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured such that a bit shaped function ($s_{i,j}(t)$) provided for a given value $b_{\text{diff}}(i, j)$ of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$, 401-40N_f) is overlapped

with a bit shaped function ($s_{i,j-1}(t)$) of a temporally preceding value ($b_{\text{diff}}(i, j-1)$) of the same frequency subband (i) like the given value ($b_{\text{diff}}(i, j)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) and with a bit shaped function ($s_{i,j+1}(t)$) of a temporally following value ($b_{i,j+1}(t)$) of the same frequency subband (i) like the given value ($b_{i,j}(t)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that a time domain waveform (2440, $s_i(t)$) provided by the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) contains an overlap between at least three temporally subsequent bit shaped functions ($s_{i,j}(t)$) of the same frequency subband (i).

10

3. The watermark signal provider (2400; 307) according to claim 1, wherein the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured such that a temporal extension of a bit shaping function (2450, $g_i(t)$) is a temporal range, in which the bit shaping function (2450, $g_i(t)$) comprises non zero values, and wherein the temporal range is at least three bit intervals (j) long.

15

4. The watermark signal provider (2400; 307) according to claim 1, wherein the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured such that a bit shaping function (2450, $g_i(t)$) is based on an amplitude modulated periodic signal;

20

wherein an amplitude modulation of the amplitude modulated periodic signal is based on a baseband function ($g_i^T(t)$);

25

wherein the temporal extension of the bit shaping function (2450, $g_i(t)$) is based on the baseband function ($g_i^T(t)$); and

wherein i designates an index for a frequency subband, T designates transmitter, and t designates a temporal variable.

30

5. The watermark signal provider (2400; 307) according to claim 4, wherein the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured, such that the baseband function ($g_i^T(t)$) is identical for a plurality of frequency subbands (i) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f).

35

6. The watermark signal provider (2400; 307) according to claim 4, wherein a periodic part of the bit shaping function (2450, $g_i(t)$) is based on a cosine function such that $g_i(t) = g_i^T(t) \cdot \cos(2\pi f_i t)$, wherein \cos is a cosine function and f_i is a center frequency of a corresponding frequency subband (i) of the bit shaping function (2450, $g_i(t)$).
7. The watermark signal provider(2400; 307) according to claim 1, further comprising a weight tuner (102), to tune a weight (105, $\gamma(i, j)$) of a bit shaped function ($s_{i,j}(t)$) provided for a given value ($b_{\text{diff}}(i, j)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that $s_{i,j}(t) = b_{\text{diff}}(i, j) \gamma(i, j) \cdot g_i(t - j \cdot T_b)$, wherein the weight tuner (102) is configured to tune the weight (105, $\gamma(i, j)$) such that an energy of the bit shaped function ($s_{i,j}(t)$) is maximized in regards of inaudibility.
8. The watermark signal provider (2400; 307) according to claim 1, wherein the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured such that a time domain waveform (2440, $s_i(t)$) of a given frequency subband (i) is a sum of all bit shaped functions ($s_{i,j}(t)$) of the given frequency subband (i), such that $s_i(t) = \sum_j s_{i,j}(t)$.
9. The watermark signal provider (2400; 307) according to claim 1, wherein the time domain waveform combiner (2460) is configured such that the watermark signal (2420, $wms(t)$; 307a; 101b) is a sum of the provided waveforms (2440, $s_i(t)$) for the plurality of frequency subbands (i), such that $wms(t) = \sum_i s_i(t)$.
10. A method (2500) for providing a watermark signal (2420, $wms(t)$; 307a; 101b) in dependence on a time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of watermark data, in which the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) comprises values associated to frequency subbands (i) and bit intervals (j), the method (2500)-comprising:
- providing (2510) time domain waveforms (2440, $s_i(t)$) for a plurality of frequency subbands (i), based on the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the watermark data, by mapping a given value ($b_{\text{diff}}(i, j)$) of the time frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) onto a bit shaping

function (2450, $g_i(t)$), wherein a temporal extension of the bit shaping function (2450, $g_i(t)$) is longer than the bit interval (j) associated to the given value ($b_{\text{diff}}(i, j)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that there is a temporal overlap between bit shaped functions ($s_{i,j}(t)$) provided for temporally subsequent values of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency subband (i), and such that a time domain waveform (2440, $s_i(t)$) of a given frequency subband (i) contains a plurality of bit shaped functions ($s_{i,j}(t)$) provided for temporally subsequent values of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency band (i); and

combining (2520) the provided time-domain waveforms (2440, $s_i(t)$) for the plurality of frequencies to derive the watermark signal (2420, wms(t); 307a; 101b).

11. A computer program for performing the method according to claim 10 when the computer program runs on a computer.

12. A watermark signal provider (2400; 307) for providing a watermark signal (2420, wms(t); 307a; 101b) in dependence on a time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of watermark data, in which the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) comprises values associated to frequency subbands (i) and bit intervals (j), the watermark signal provider (2400; 307) comprising:

a time-frequency-domain waveform provider (2430; 411-41N_f, 421-42N_f) configured to provide time-domain waveforms (2440; $s_i(t)$) for a plurality of frequency subbands (i), based on the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the watermark data, wherein the time-frequency-domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured to map a given value ($b_{\text{diff}}(i, j)$) of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) onto a bit shaping function ($g_i(t)$), wherein a temporal extension of the bit shaping function ($g_i(t)$) is longer than the bit interval (j) associated to the given value ($b_{\text{diff}}(i, j)$) of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that there is a temporal overlap between bit shaped functions ($g_i(t)$) provided for temporally subsequent values of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency subband (i); and

wherein the time-frequency-domain waveform provider (2430; 411-41N_f, 421-42N_f) is further configured such that a time-domain waveform (2440, $s_i(t)$) of a given frequency subband (i) contains a plurality of bit shaped functions ($s_{i,j}(t)$) provided for temporally subsequent values of the time-frequency-domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency band (i); and

a time-domain waveform combiner (2460), to combine the provided time-domain waveforms (2440, $s_i(t)$) for the plurality of frequencies (i) of the time-frequency-domain provider (2430; 411-41N_f, 421-42N_f) to derive the watermark signal (2420, wms(t); 307a; 101b);

wherein the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) is configured such that a bit shaped function ($s_{i,j}(t)$) provided for a given value $b_{\text{diff}}(i, j)$ of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$, 401-40N_f) is overlapped with a bit shaped function ($s_{i,j-1}(t)$) of a temporally preceding value ($b_{\text{diff}}(i, j-1)$) of the same frequency subband (i) like the given value ($b_{\text{diff}}(i, j)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) and with a bit shaped function ($s_{i,j+1}(t)$) of a temporally following value ($b_{i,j+1}(t)$) of the same frequency subband (i) like the given value ($b_{i,j}(t)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that a time domain waveform (2440, $s_i(t)$) provided by the time-frequency domain waveform provider (2430; 411-41N_f, 421-42N_f) contains an overlap between at least three temporally subsequent bit shaped functions ($s_{i,j}(t)$) of the same frequency subband (i).

13. A method (2500) for providing a watermark signal (2420, wms(t); 307a; 101b) in dependence on a time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of watermark data, in which the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) comprises values associated to frequency subbands (i) and bit intervals (j), the method (2500) comprising:

providing (2510) time domain waveforms (2440, $s_i(t)$) for a plurality of frequency subbands (i), based on the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the watermark data, by mapping a given value ($b_{\text{diff}}(i, j)$) of the time frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) onto a bit shaping function (2450, $g_i(t)$), wherein a temporal extension of the bit shaping function

(2450, $g_i(t)$) is longer than the bit interval (j) associated to the given value ($b_{\text{diff}}(i, j)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that there is a temporal overlap between bit shaped functions ($s_{i,j}(t)$) provided for temporally subsequent values of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency subband (i), and such that a time domain waveform (2440, $s_i(t)$) of a given frequency subband (i) contains a plurality of bit shaped functions ($s_{i,j}(t)$) provided for temporally subsequent values of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) of the same frequency band (i); and

combining (2520) the provided time-domain waveforms (2440, $s_i(t)$) for the plurality of frequencies to derive the watermark signal (2420, $wms(t)$; 307a; 101b);

wherein a bit shaped function ($s_{i,j}(t)$) provided for a given value $b_{\text{diff}}(i, j)$ of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$, 401-40N_f) is overlapped with a bit shaped function ($s_{i,j-1}(t)$) of a temporally preceding value ($b_{\text{diff}}(i, j-1)$) of the same frequency subband (i) like the given value ($b_{\text{diff}}(i, j)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f) and with a bit shaped function ($s_{i,j+1}(t)$) of a temporally following value ($b_{\text{diff}}(i, j+1)$) of the same frequency subband (i) like the given value ($b_{\text{diff}}(i, j)$) of the time-frequency domain representation (2410; $b_{\text{diff}}(i, j)$; 401-40N_f), such that the provided time domain waveform (2440, $s_i(t)$) contains an overlap between at least three temporally subsequent bit shaped functions ($s_{i,j}(t)$) of the same frequency subband (i).

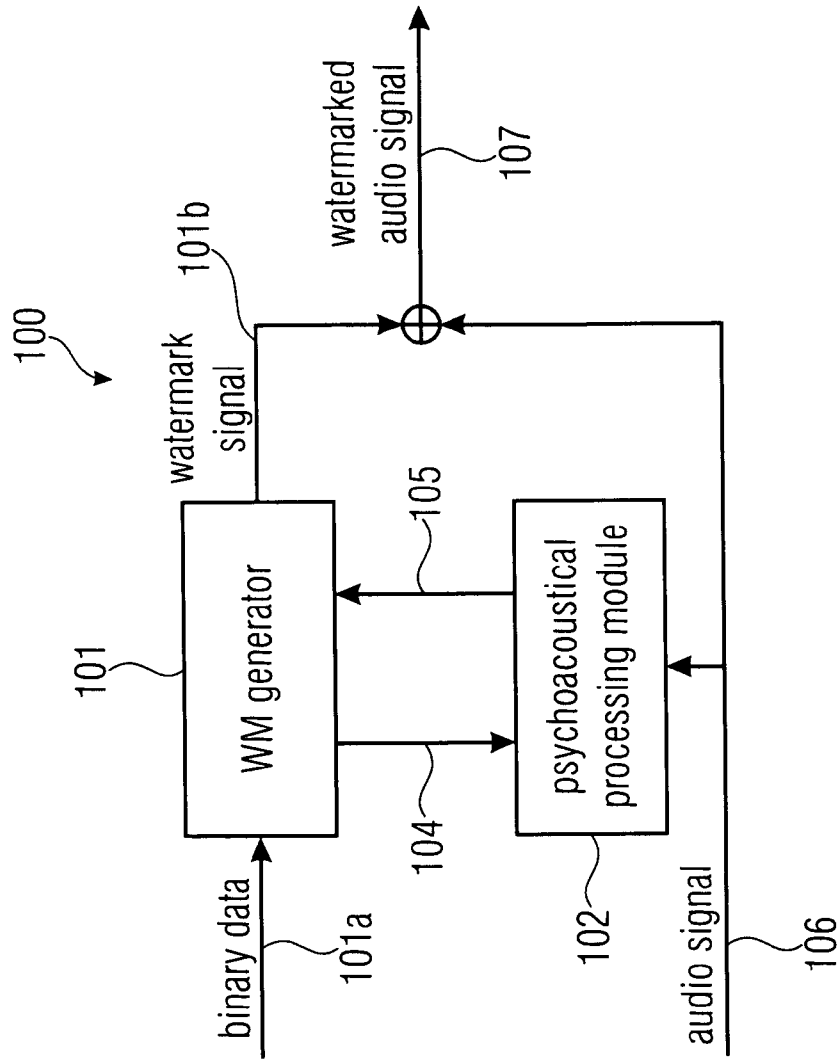


FIGURE 1

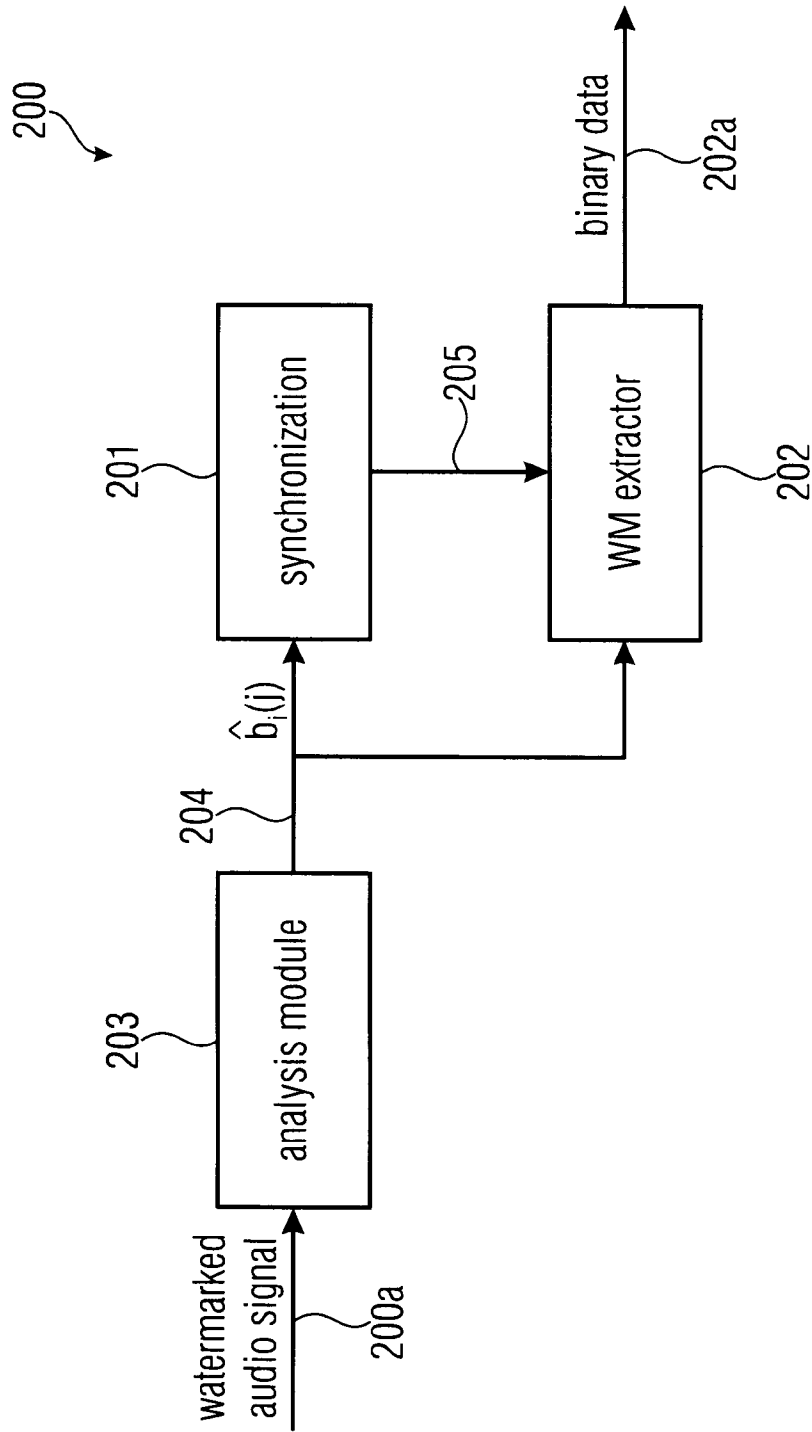


FIGURE 2

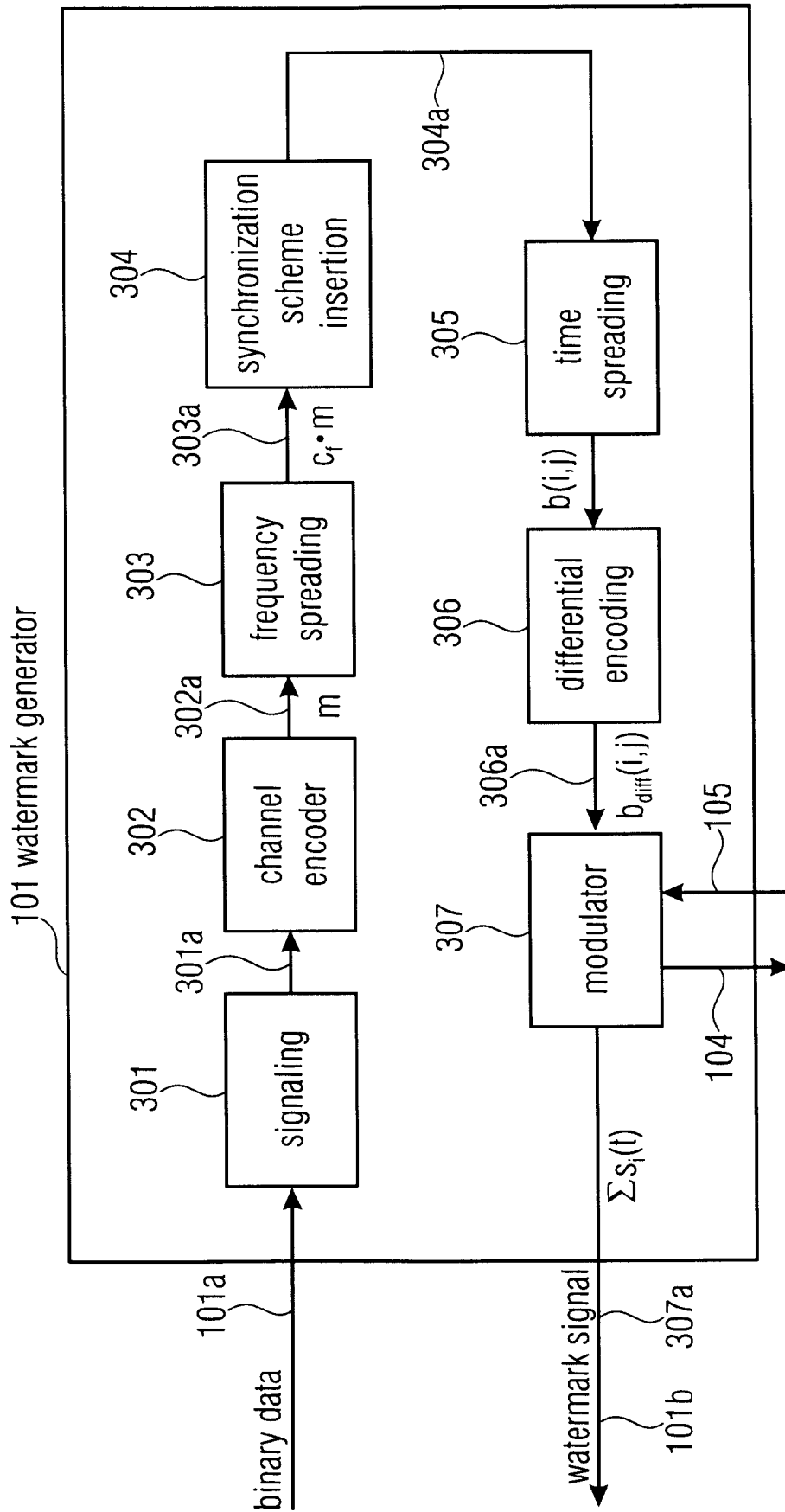


FIGURE 3

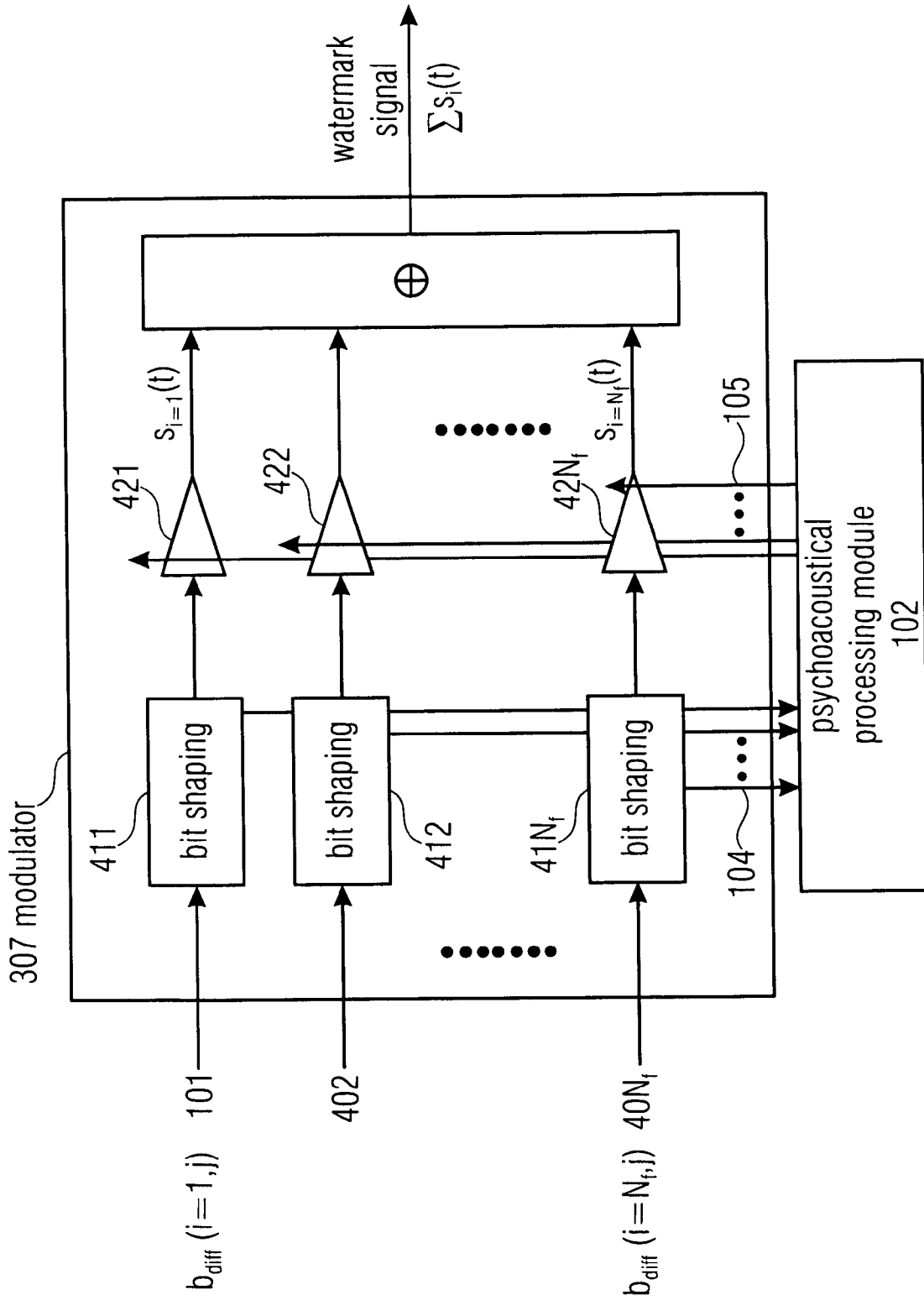


FIGURE 4

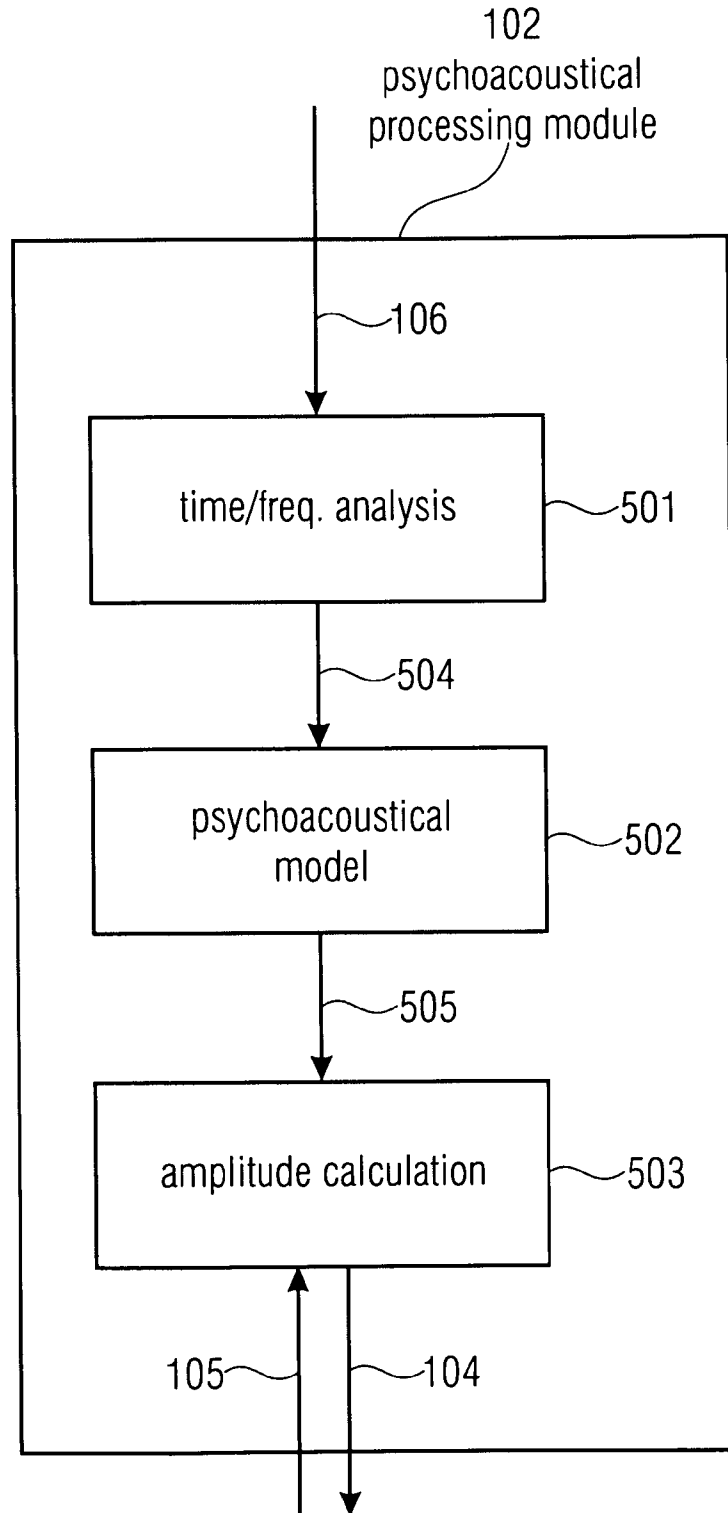


FIGURE 5

6/35

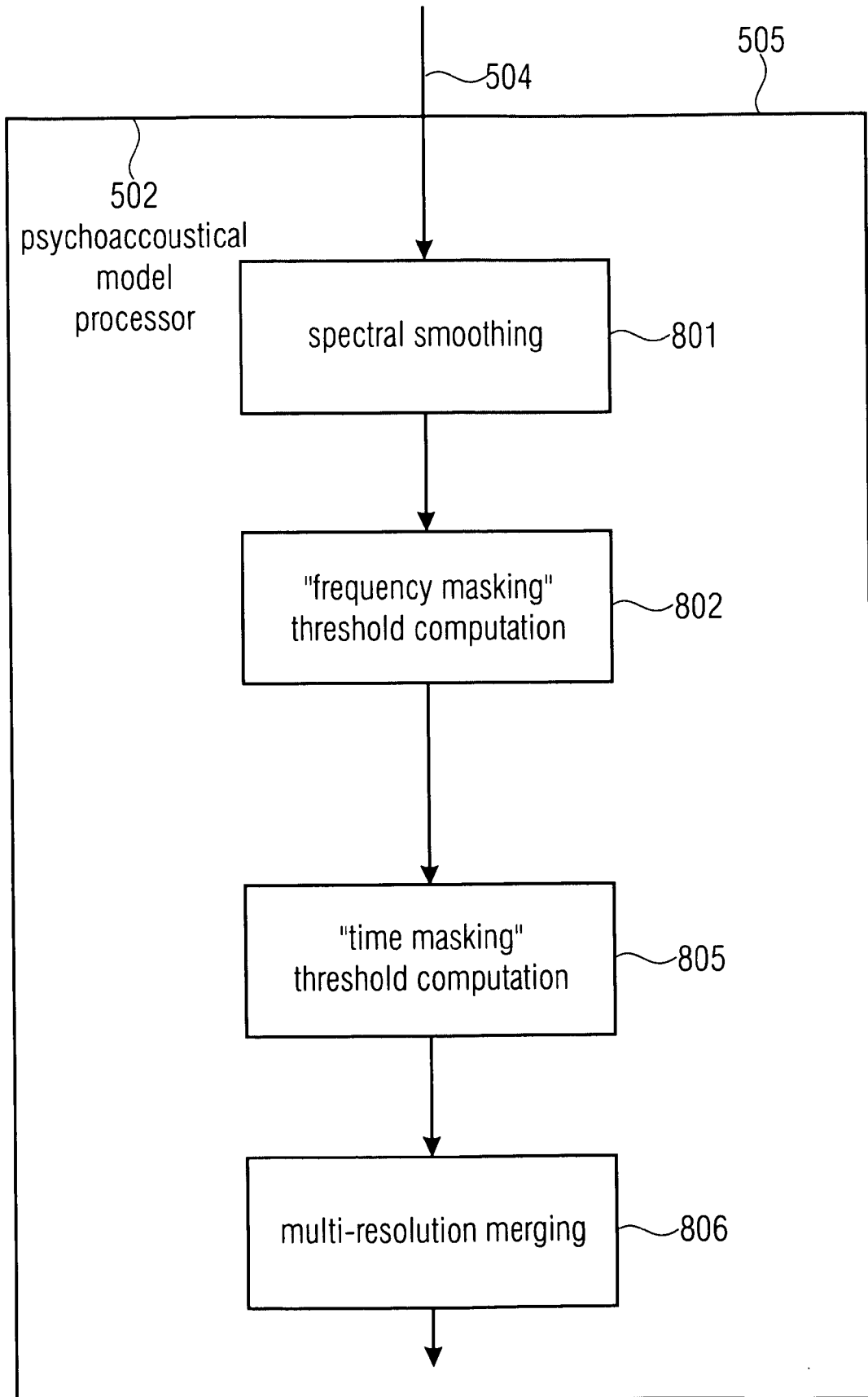


FIGURE 6

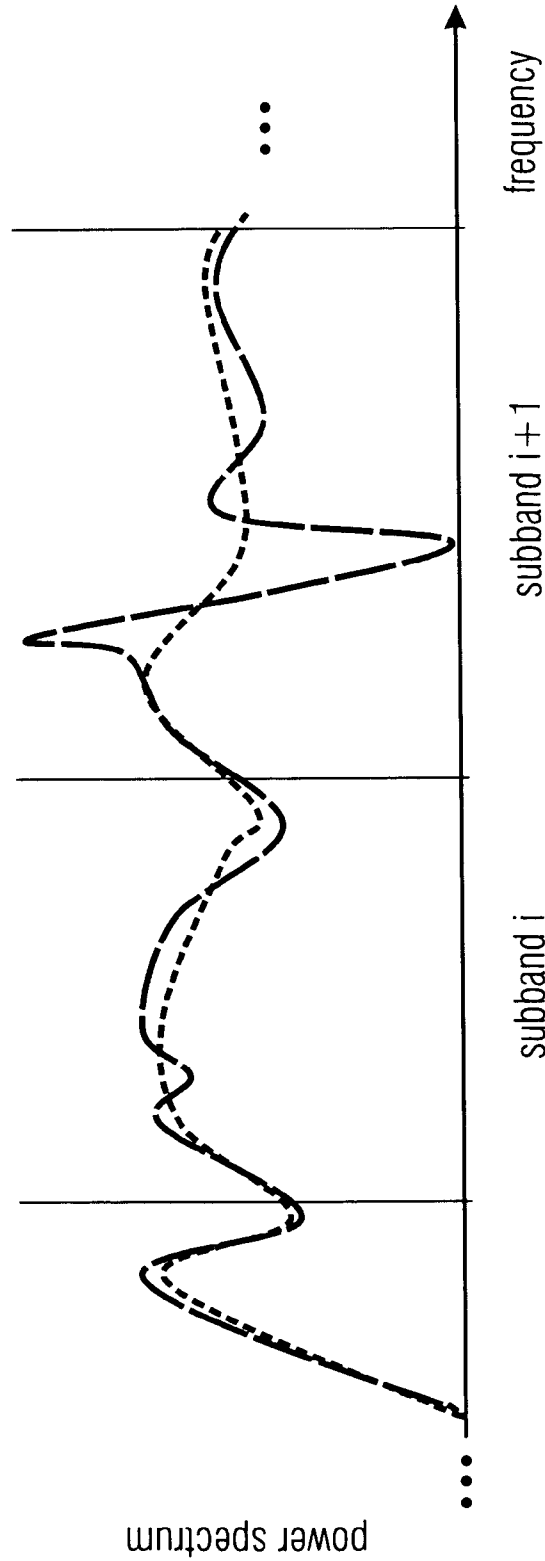


FIGURE 7

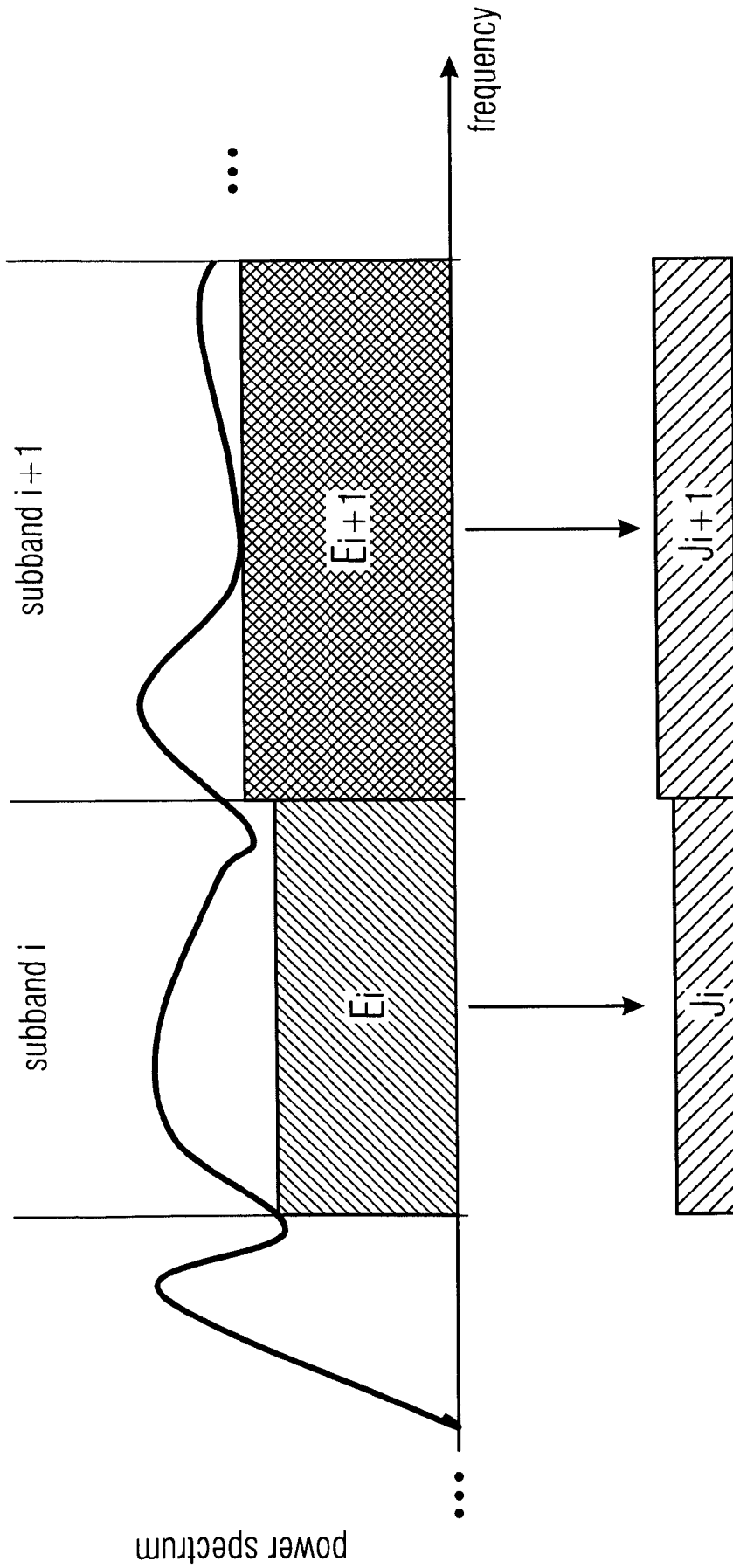


FIGURE 8

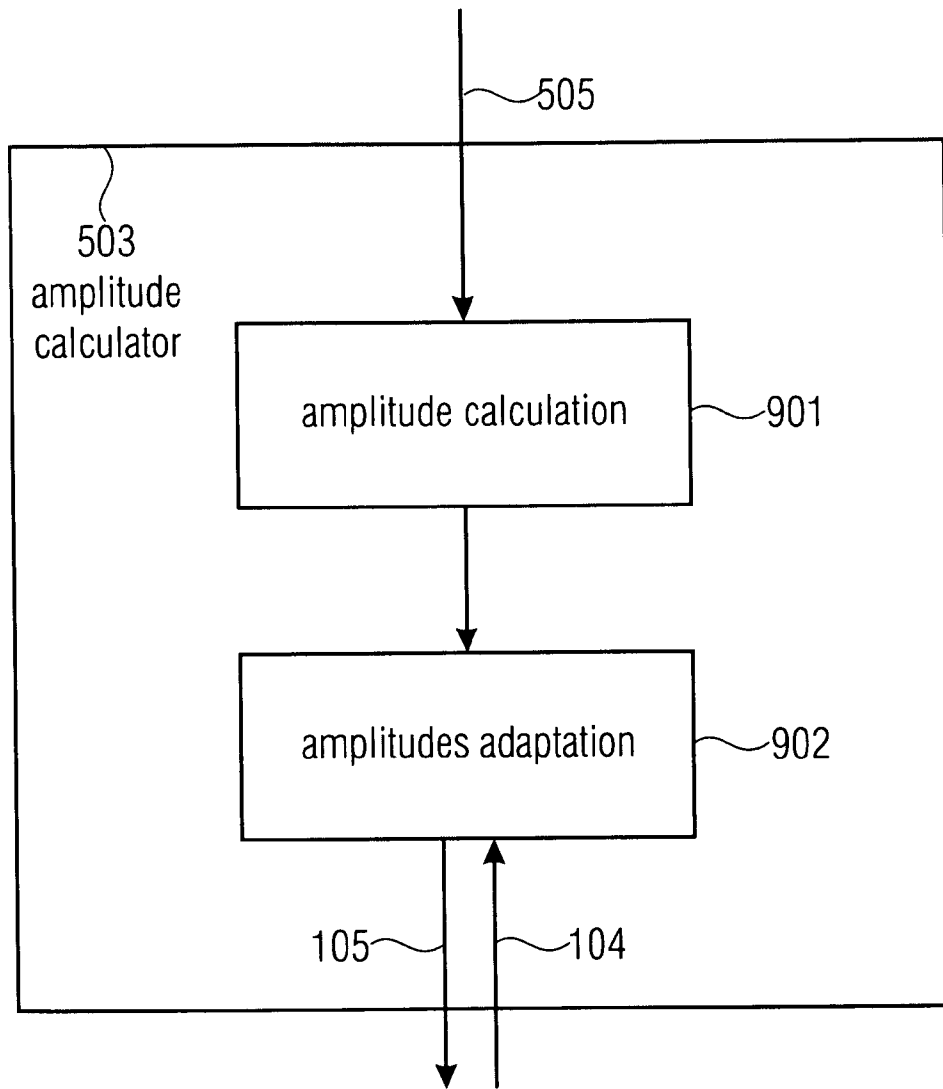


FIGURE 9

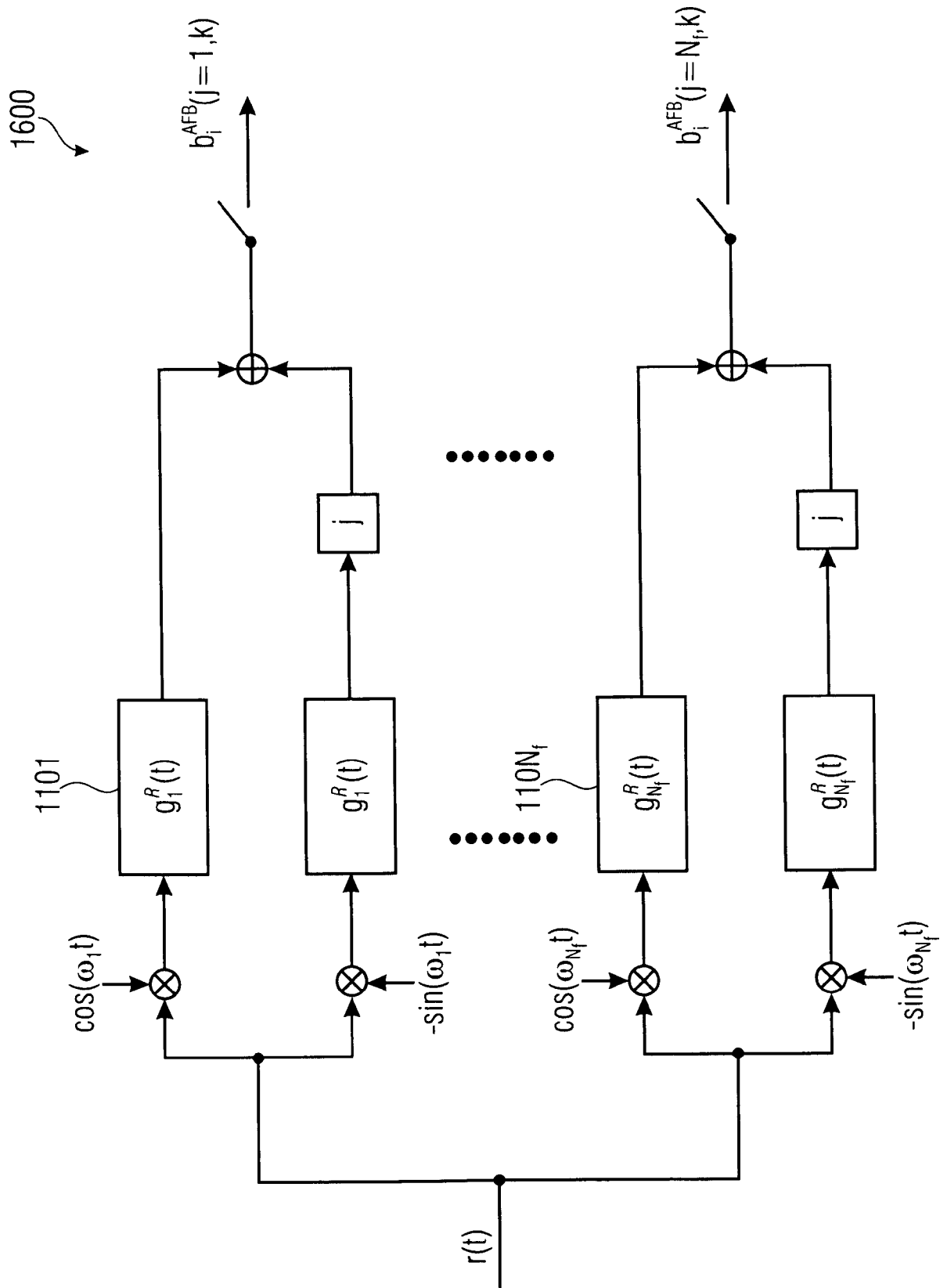


FIGURE 10A

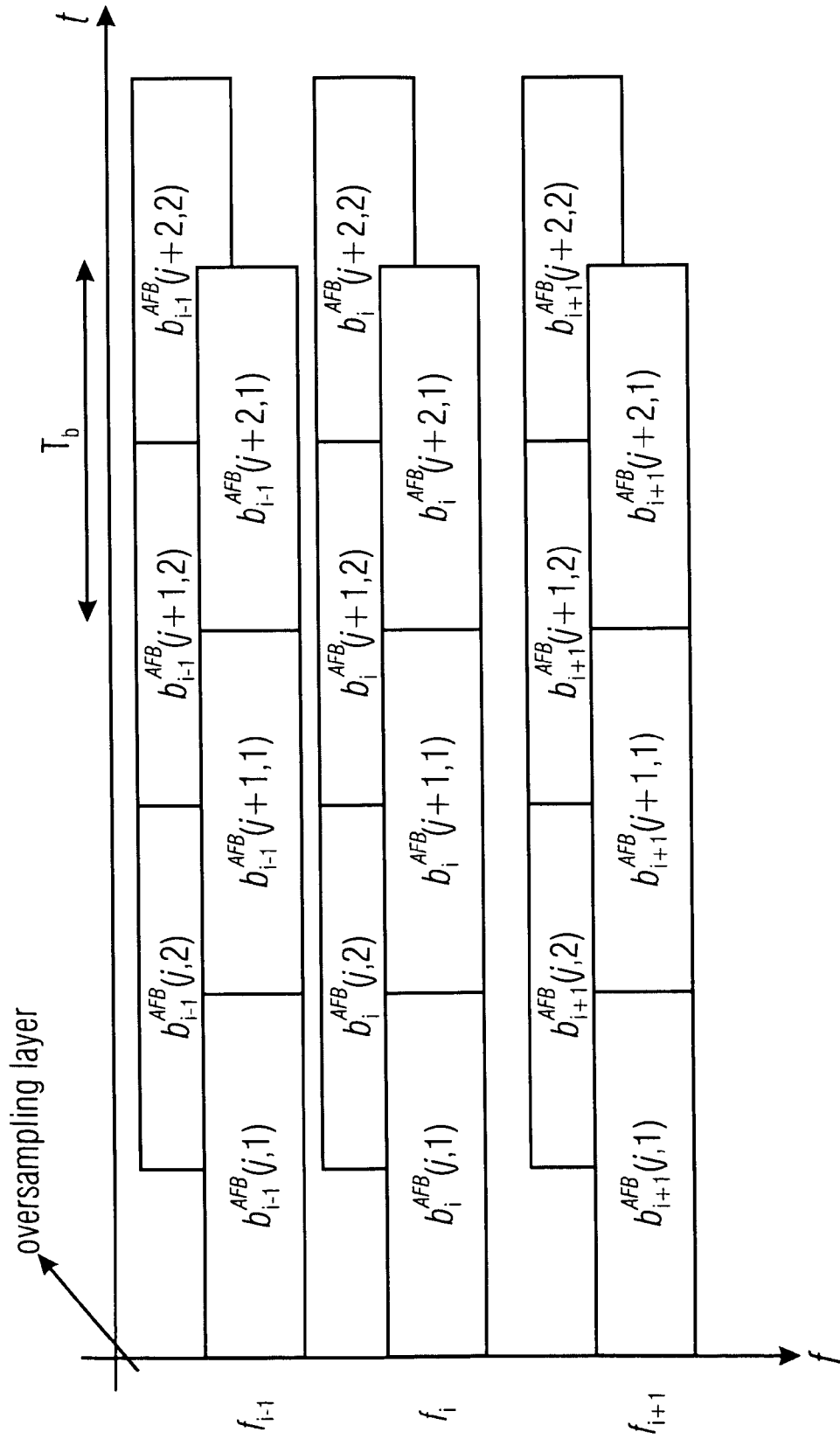


FIGURE 10B

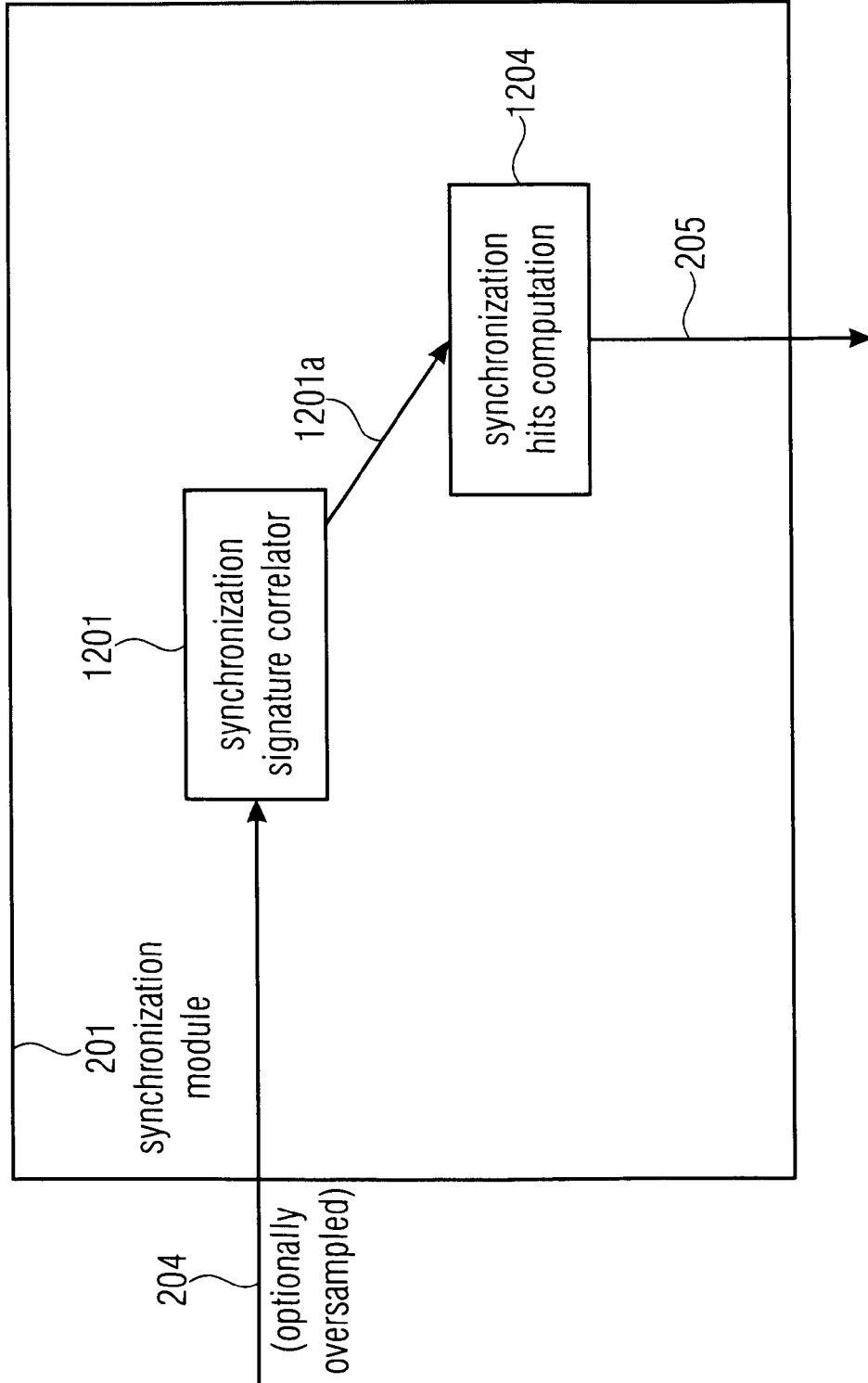


FIGURE 11A

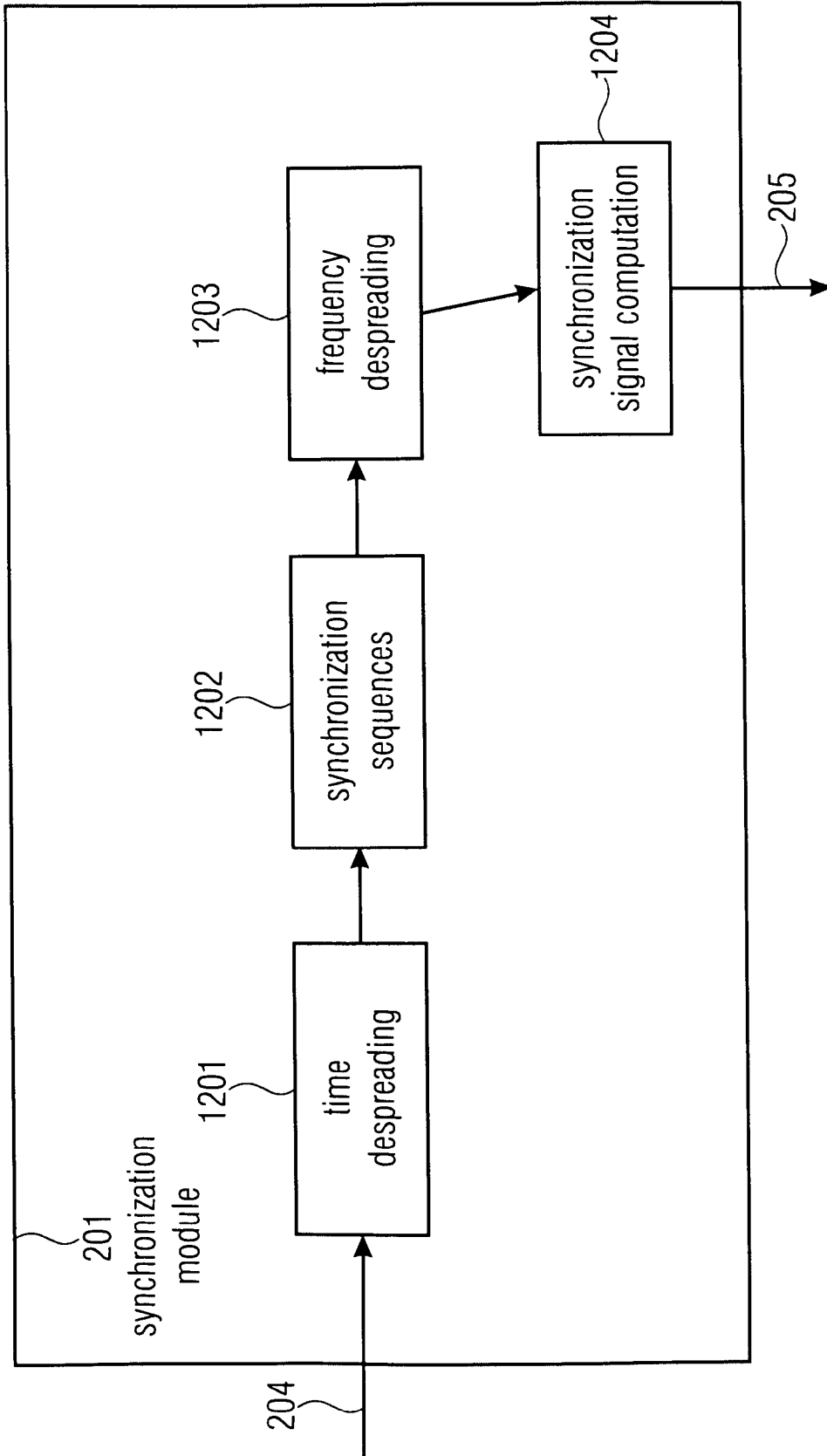


FIGURE 11B

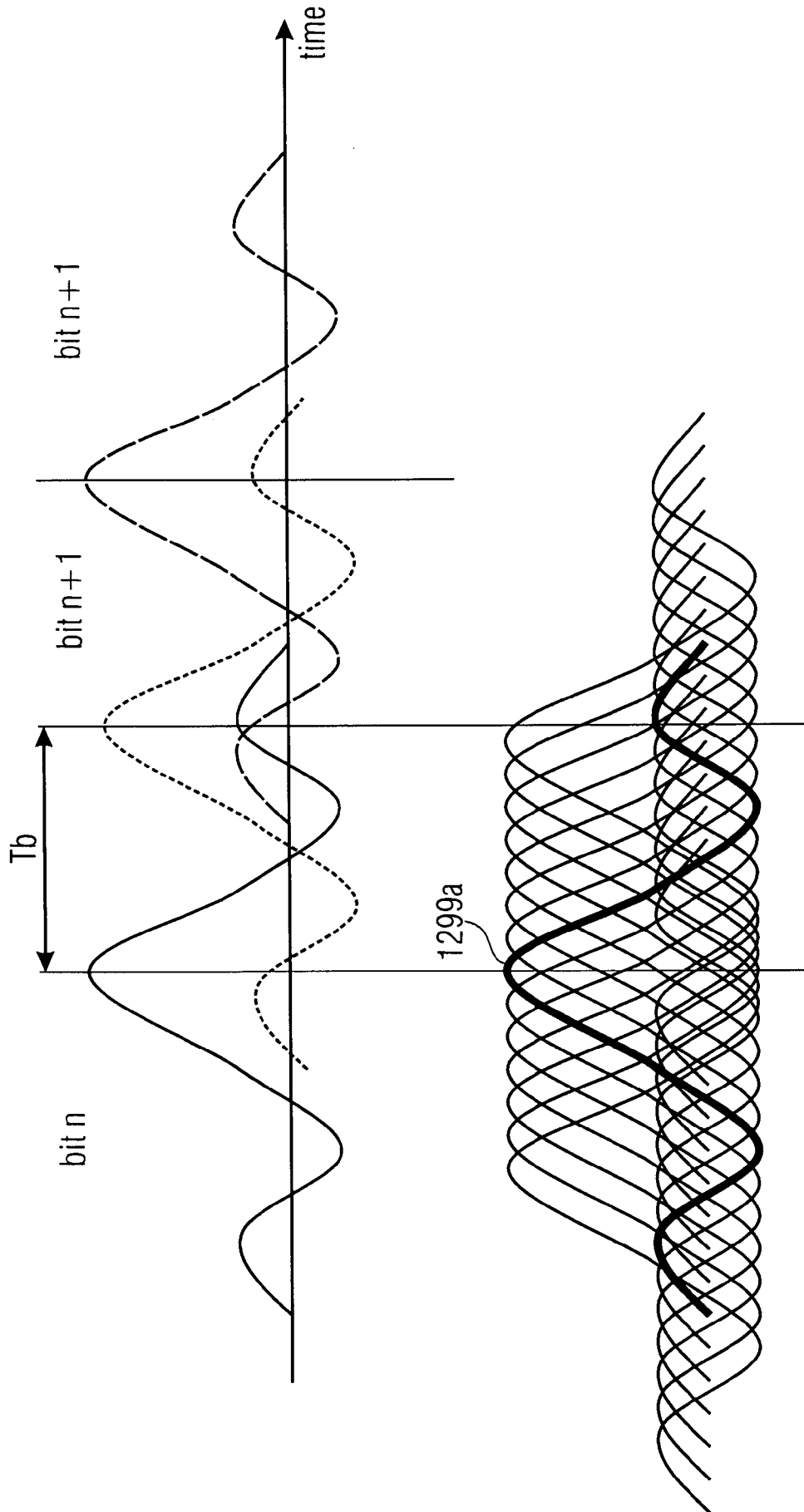


FIGURE 12A

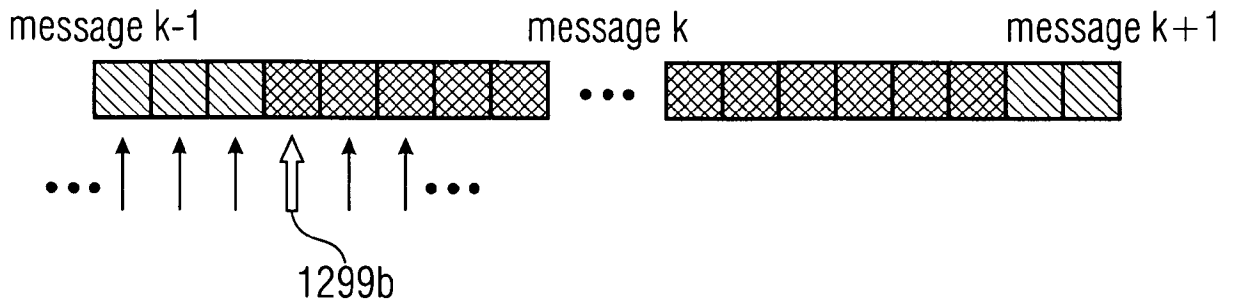


FIGURE 12B

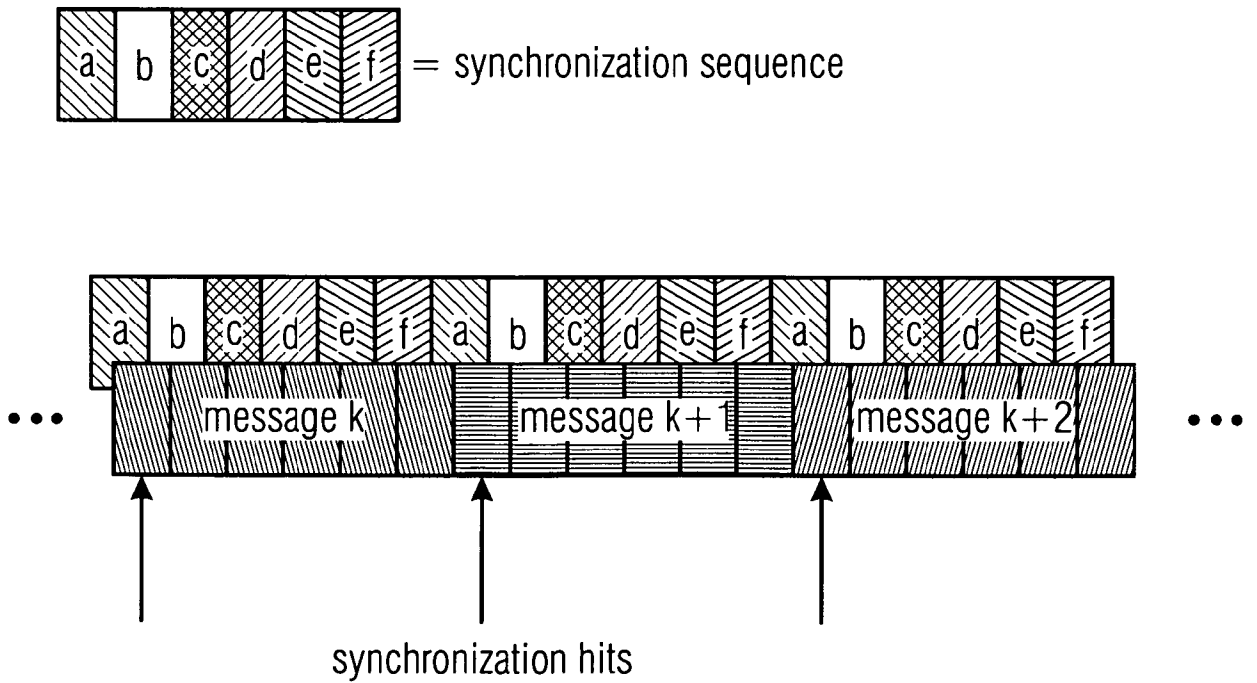


FIGURE 12C

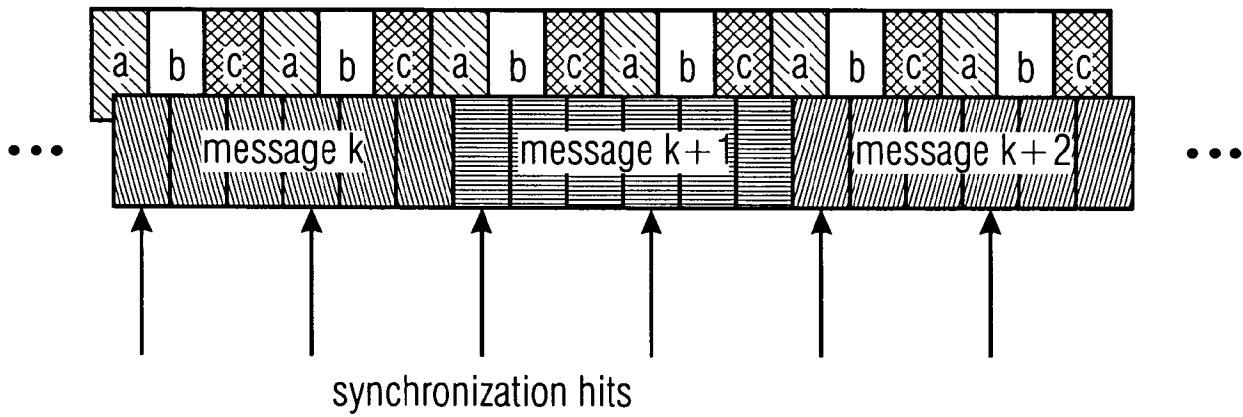
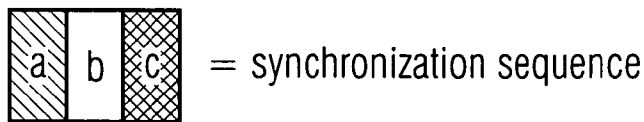


FIGURE 12D

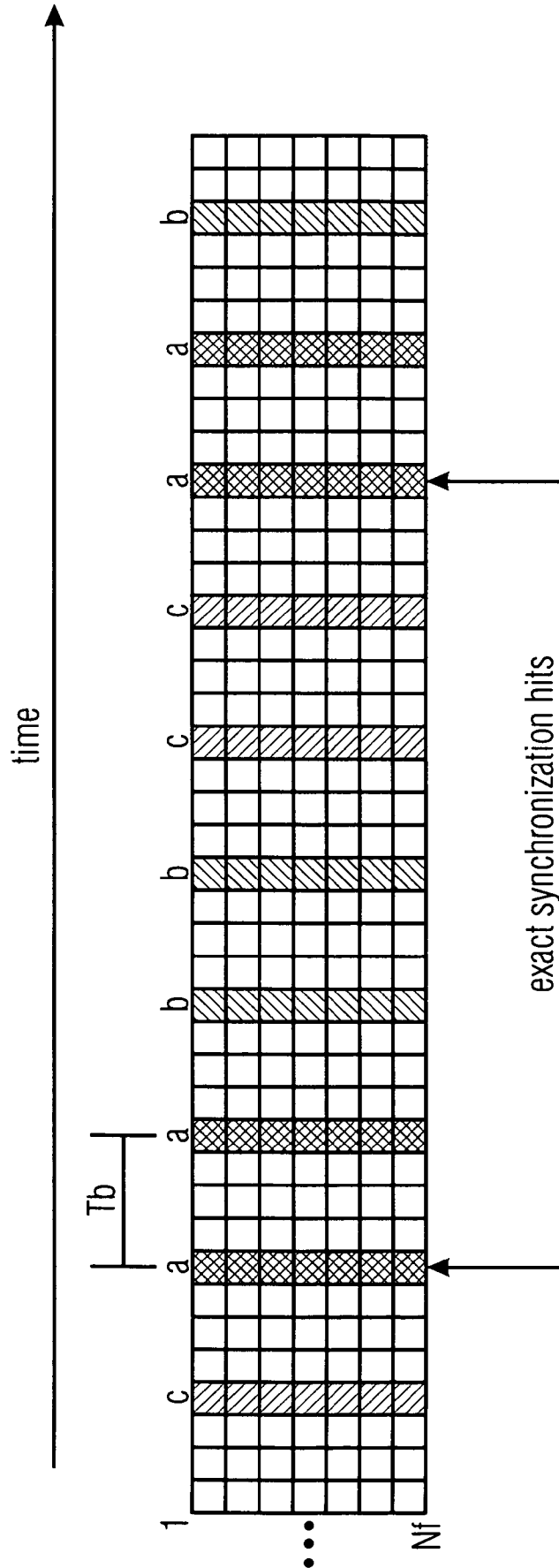


FIGURE 12E

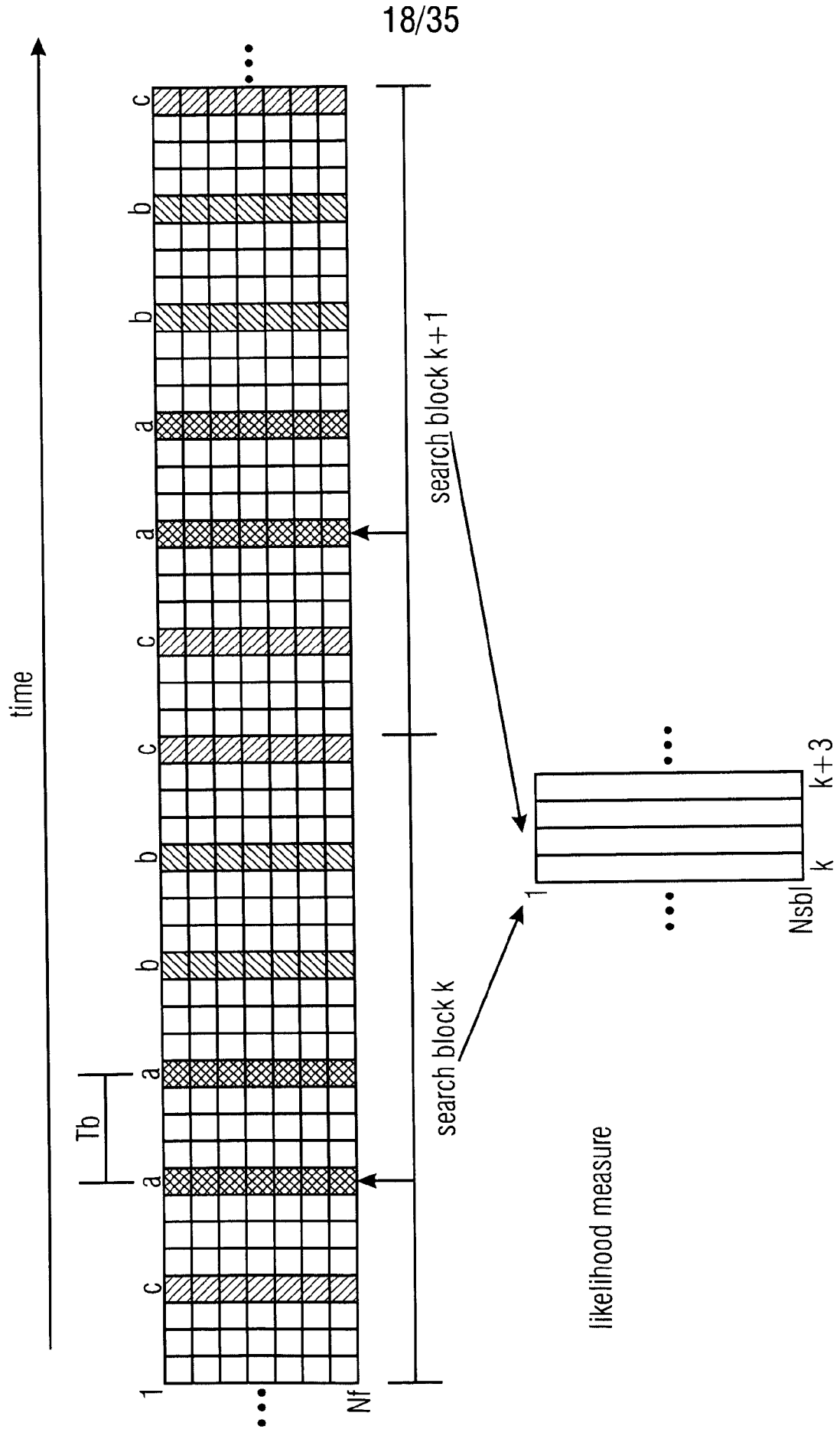


FIGURE 12F

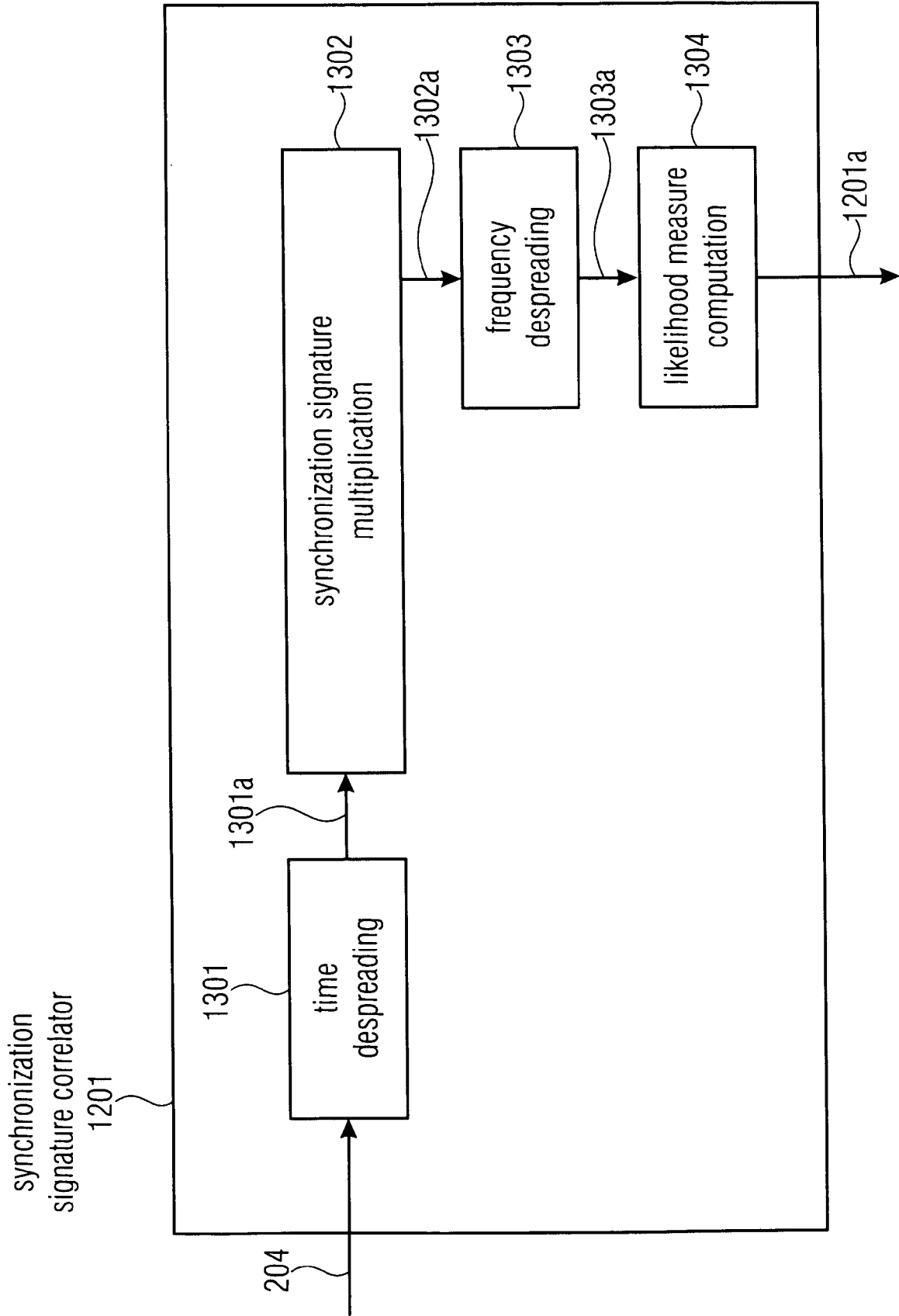


FIGURE 12G

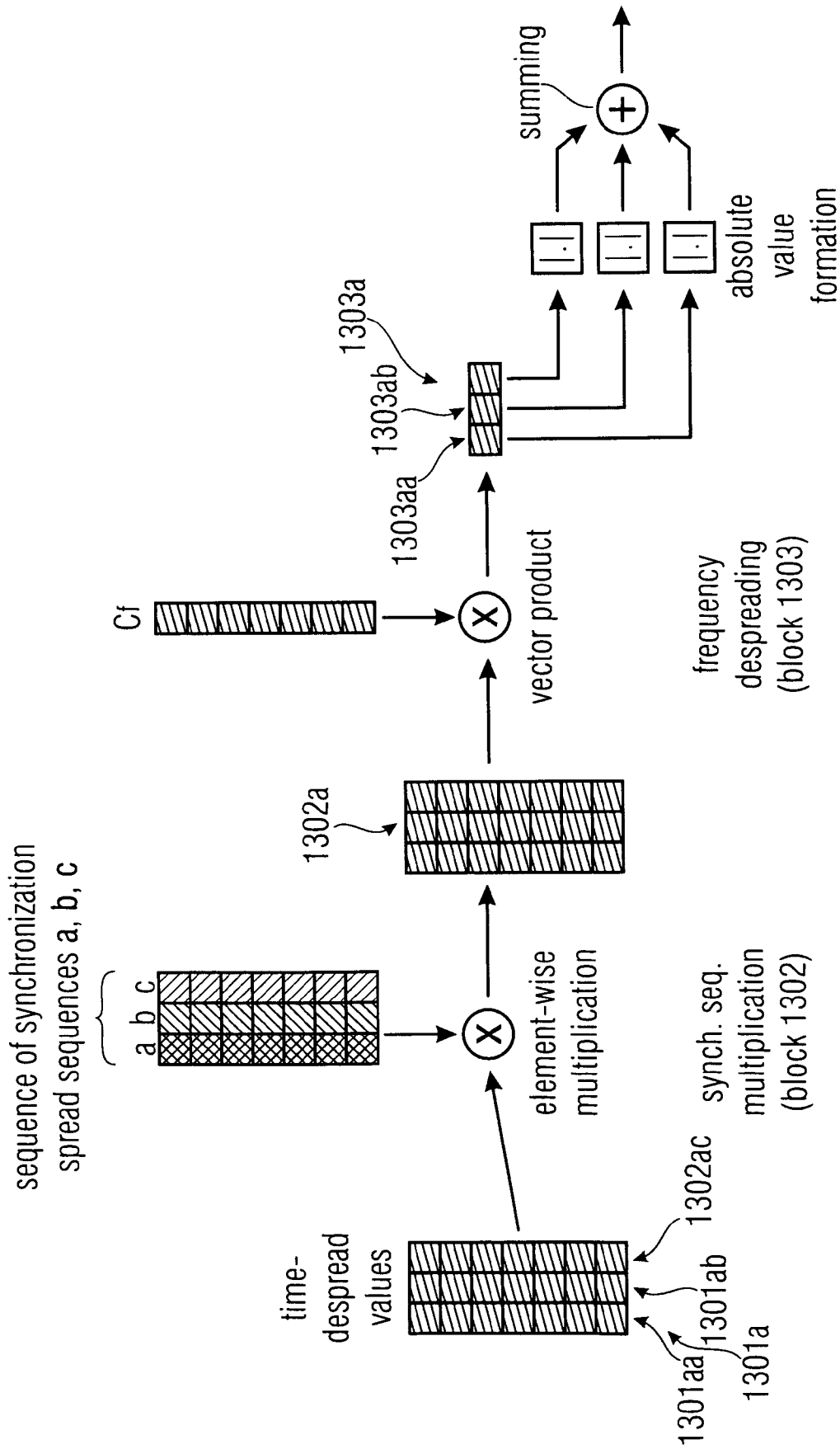


FIGURE 13B

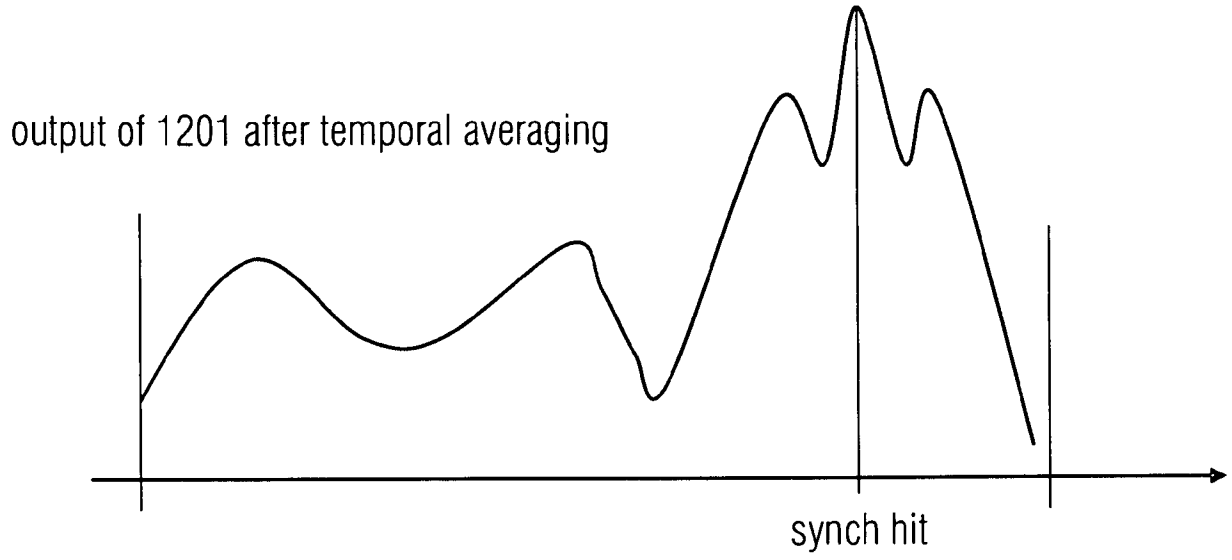


FIGURE 13C

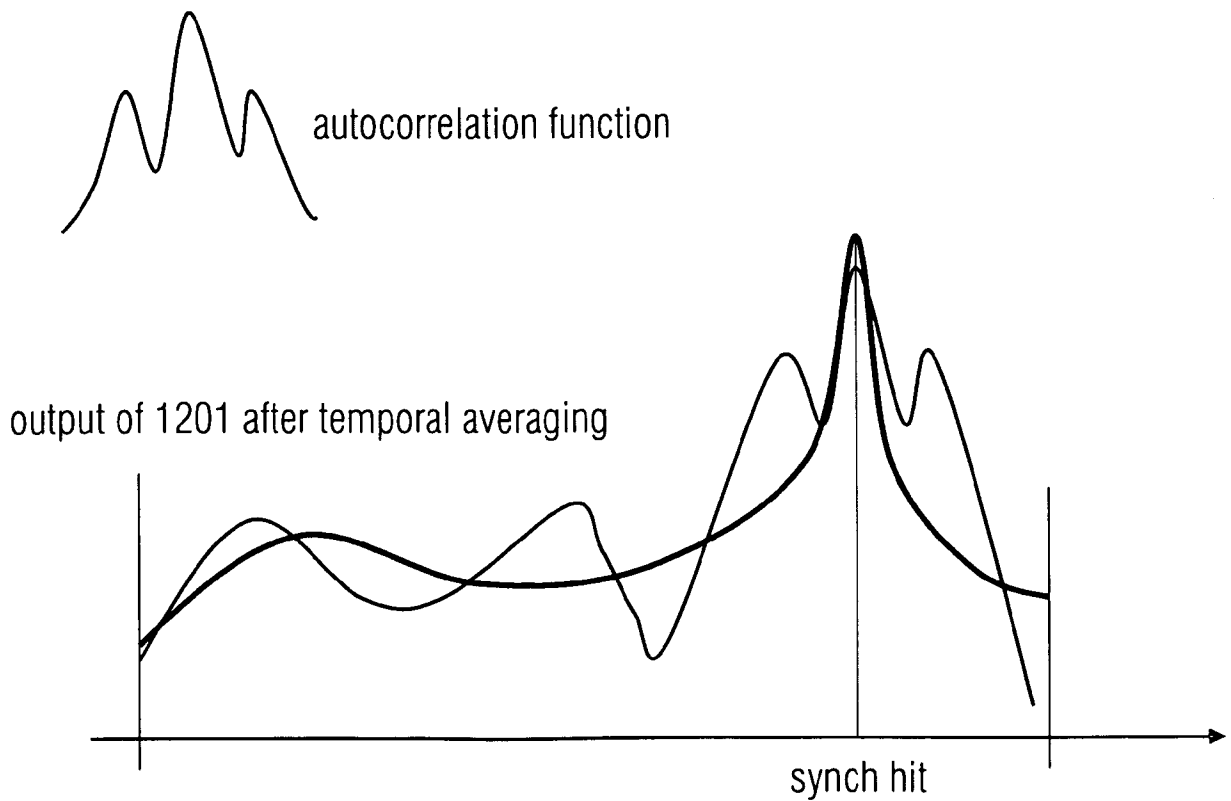


FIGURE 13D

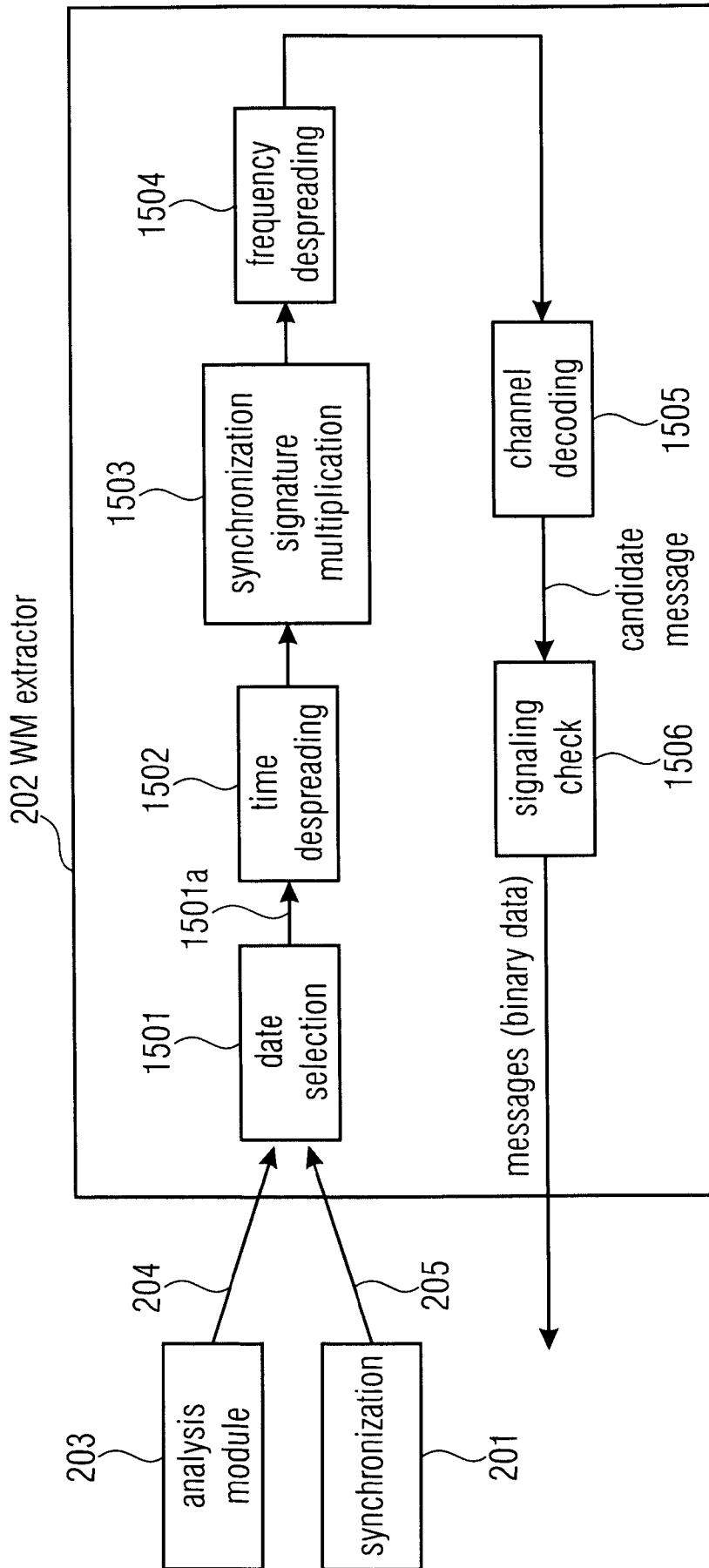


FIGURE 14

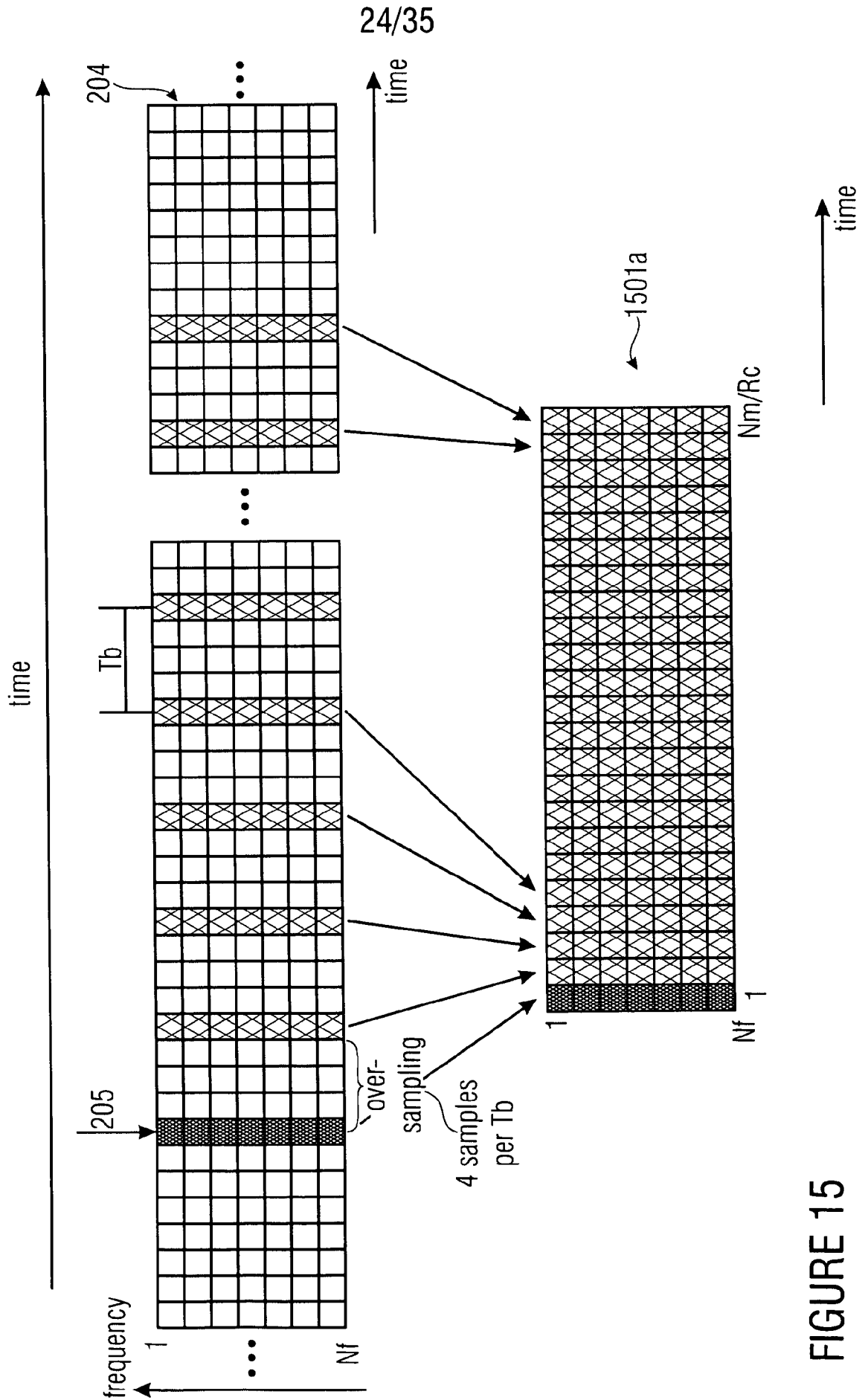


FIGURE 15

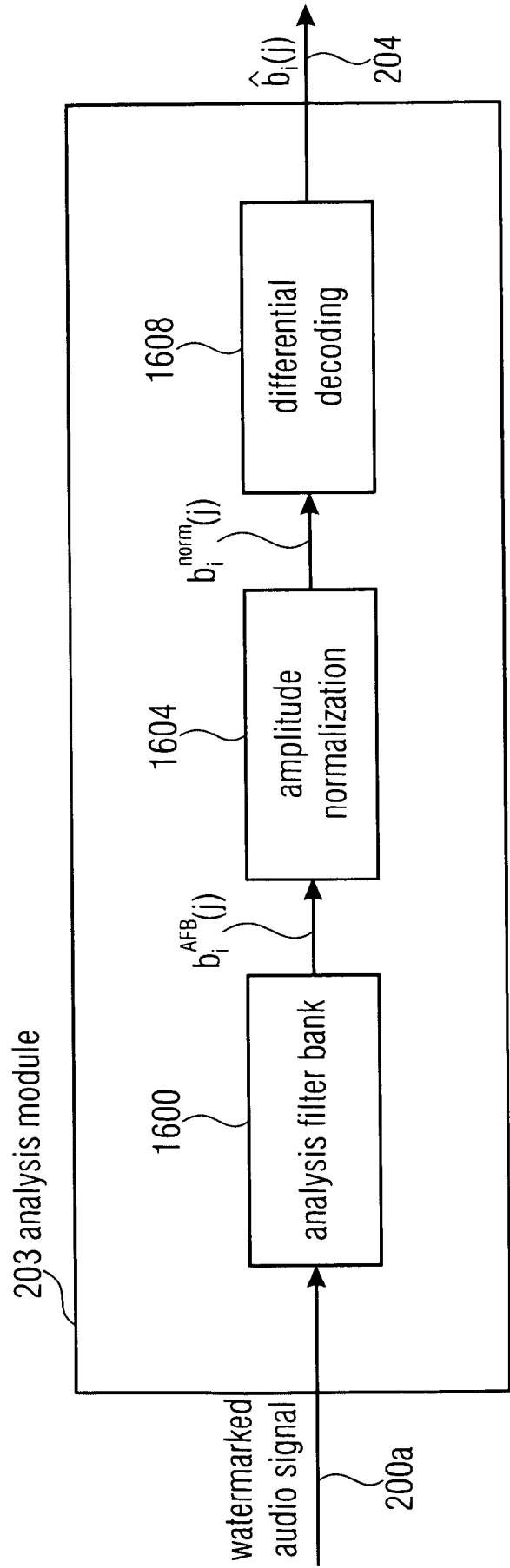


FIGURE 16



FIGURE 17A

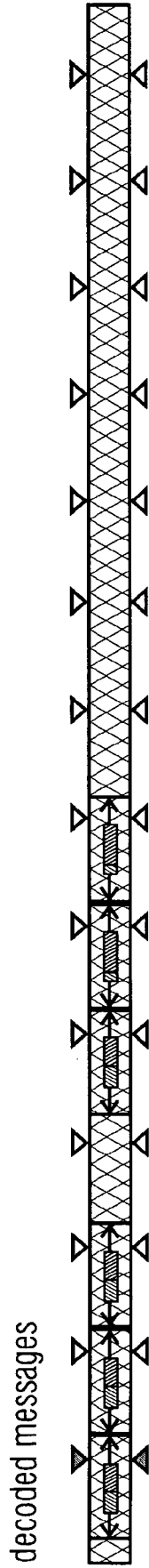


FIGURE 17B

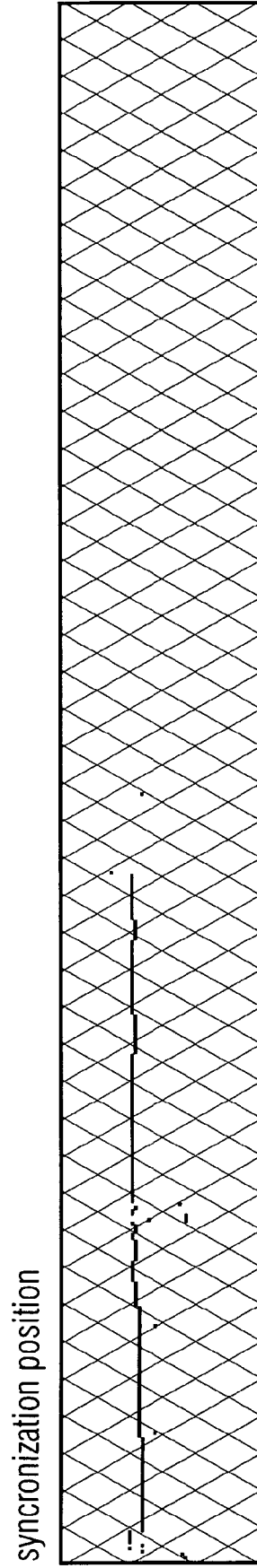


FIGURE 17C

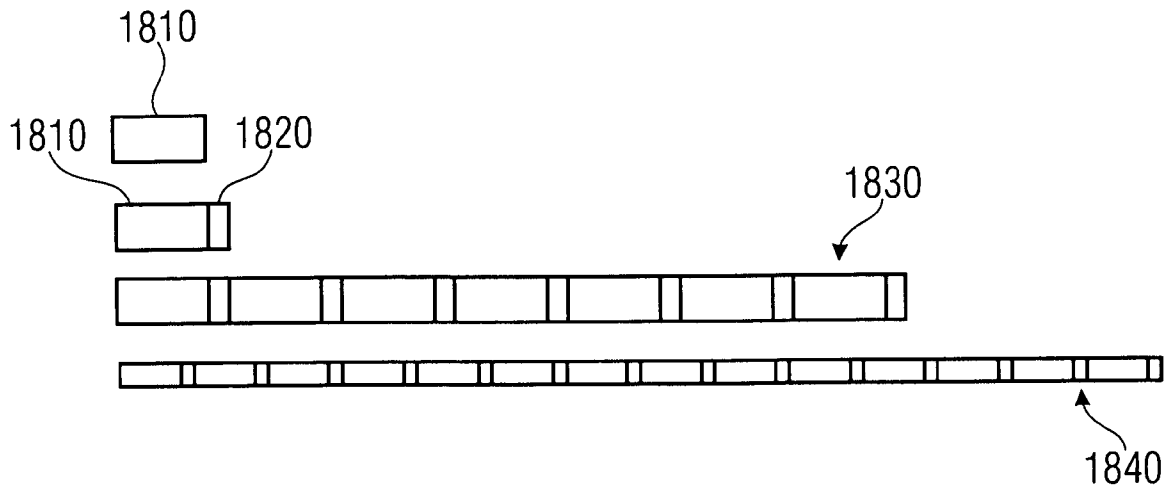


FIGURE 18A

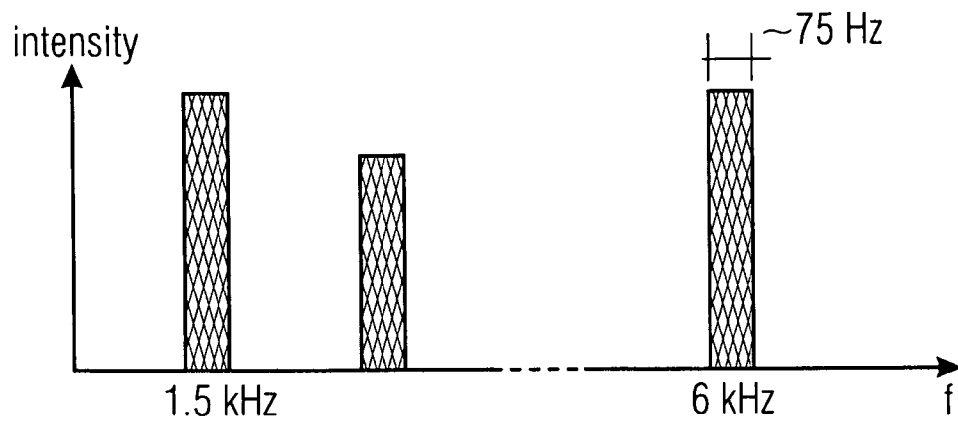


FIGURE 18B

1900

ABC synchrony - the basic concept

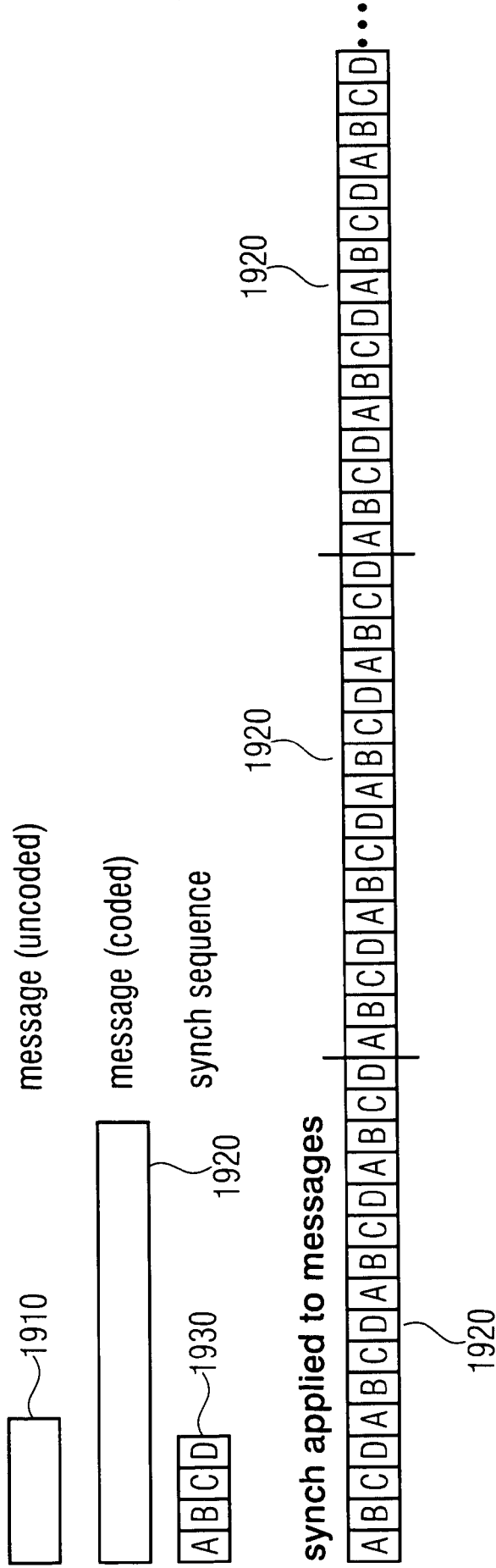
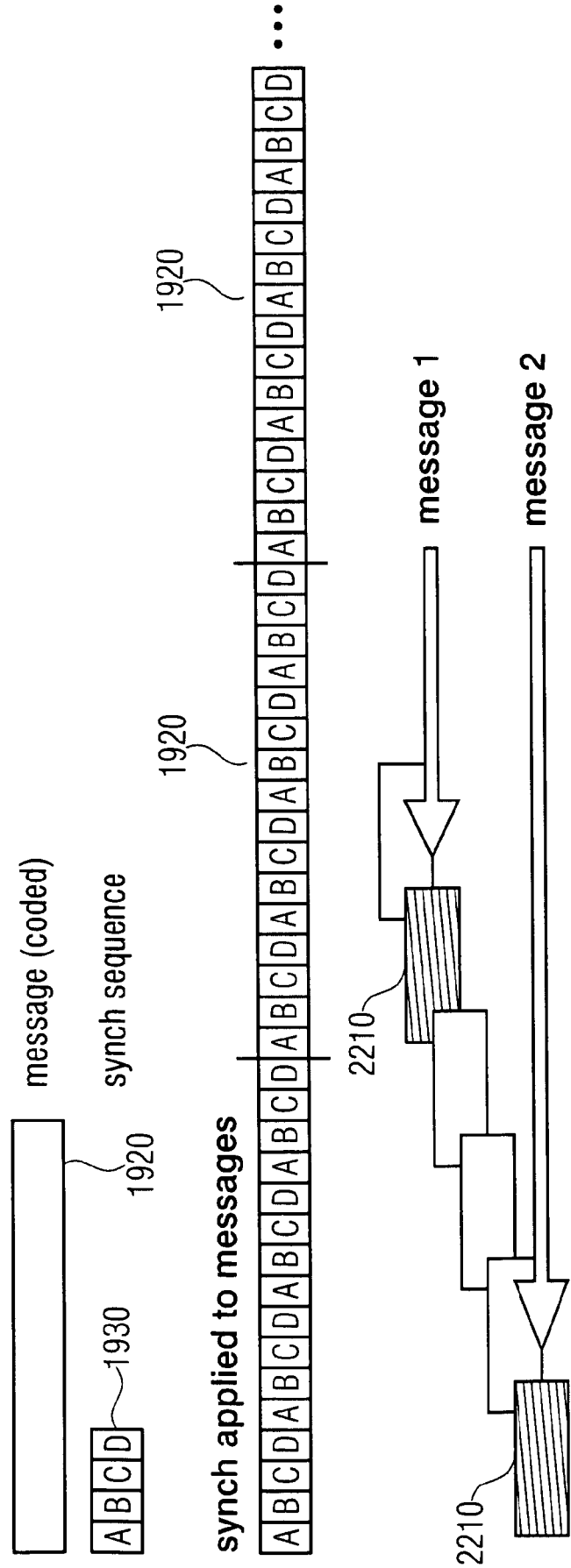


FIGURE 19

ABC synchrony - the basic concept



3. true messages are identified by means of a CRC sequences and/or a plausibility check

FIGURE 22

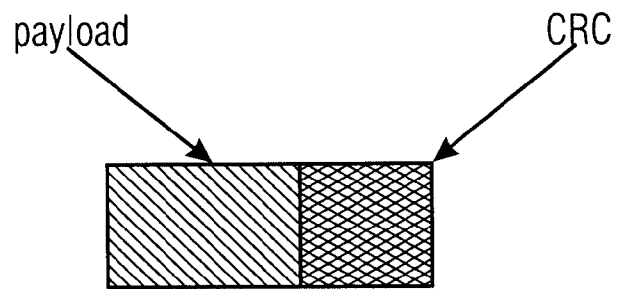


FIGURE 23

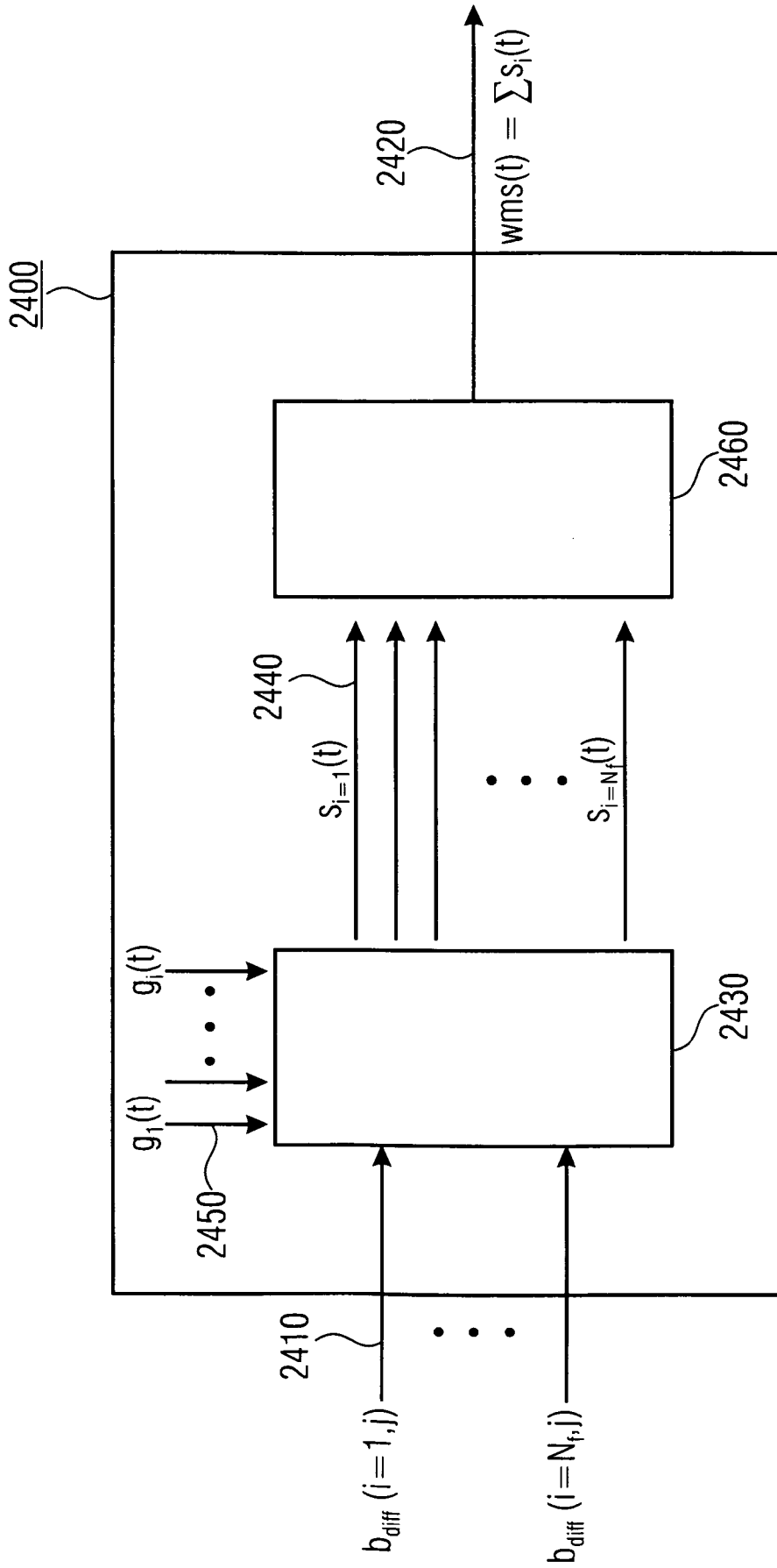


FIGURE 24

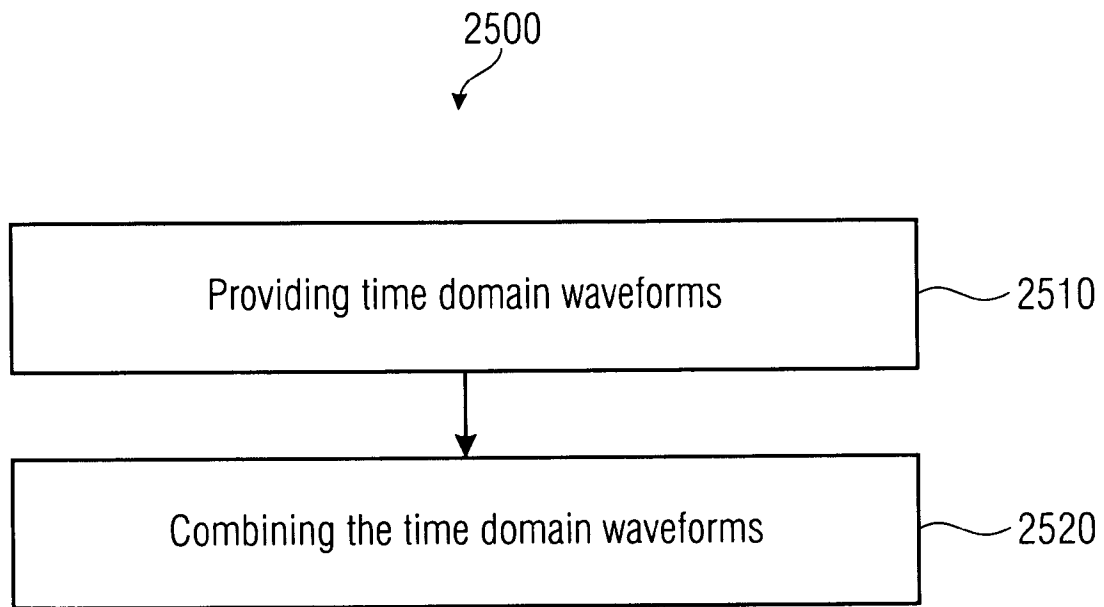


FIGURE 25

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/052694

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/00
ADD.
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)
G10L
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 practical, search terms used)
EPO-Internal, INSPEC, WPI Data, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KIROVSKI D ET AL: "Robust spread-spectrum audio watermarking", 2001 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). SALT LAKE CITY, UT, vol. 3, 7 May 2001 (2001-05-07), pages 1345-1348, XP010803141, [IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP)], NEW YORK, NY : IEEE, US ISBN: 978-0-7803-7041-8	1,3-11
A	abstract page 1346, paragraph [3.1 Mechanism against Desynchronization Attacks] figure 1 ----- -/--	2,12,13

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
- "E" earlier document but published on or after the international filing date
- "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)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "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
- "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
- "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.
- "&" document member of the same patent family

Date of the actual completion of the international search 23 May 2011	Date of mailing of the international search report 30/05/2011
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Greiser, Norbert

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/052694

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>KIROVSKI D, MALVAR H: "Robust Covert communication over a Public Audio Channel using Spread Spectrum", LECTURE NOTES IN COMPUTER SCIENCE, [Online] vol. 2137/2001, 1 January 2001 (2001-01-01), pages 354-368, XP002590801, Berlin/Heidelberg Retrieved from the Internet: URL:http://www.springerlink.com/content/20e7cu8c861fmw9m/fulltext.pdf [retrieved on 2010-07-07] abstract page 363, paragraph 3 - page 364, paragraph 1 figure 4</p> <p style="text-align: center;">-----</p>	1,10,11
A	<p>DE 10 2008 014311 A1 (FRAUNHOFER GES FORSCHUNG [DE]) 17 September 2009 (2009-09-17) abstract</p> <p style="text-align: center;">-----</p>	1,10-13

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2011/052694

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
DE 102008014311 A1	17-09-2009	CN 102007530 A EP 2266114 A1 WO 2009112184 A1	06-04-2011 29-12-2010 17-09-2009
