US007382886B2

(12) **United States Patent**
Henn et al.

(10) **Patent No.:** US 7,382,886 B2
(45) **Date of Patent:** Jun. 3, 2008

(54) **EFFICIENT AND SCALABLE PARAMETRIC STEREO CODING FOR LOW BITRATE AUDIO CODING APPLICATIONS**

(75) Inventors: **Fredrik Henn**, Bromma (SE); **Kristofer Kjorling**, Solna (SE); **Lars Liljeryd**, Solna (SE); **Jonas Roden**, Solna (SE); **Jonas Engdegard**, Stockholm (SE)

(73) Assignee: **Coding Technologies AB**, Stockholm (SE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 593 days.

(21) Appl. No.: **10/483,453**

(22) PCT Filed: **Jul. 10, 2002**

(86) PCT No.: **PCT/SE02/01372**

§ 371 (c)(1),
(2), (4) Date: **Jan. 8, 2004**

(87) PCT Pub. No.: **WO03/007656**

PCT Pub. Date: **Jan. 23, 2003**

(51) **Int. Cl.**
H04R 5/00	(2006.01)
H04R 5/02	(2006.01)
G10L 19/00	(2006.01)
G10L 21/00	(2006.01)
H04B 1/20	(2006.01)

(52) **U.S. Cl.** **381/23**; 381/22; 381/20; 381/310; 704/500; 704/501; 369/4; 369/5

(58) **Field of Classification Search** 381/17, 381/19, 20, 21, 22, 23, 12, 310, 119, 11; 369/4–5; 704/500–501
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,463,424 A   10/1995   Dressler

(Continued)

FOREIGN PATENT DOCUMENTS

DE   19947098 A1   11/2000

(Continued)

OTHER PUBLICATIONS

Bauer, D. Examinations Regarding the Similarity of Digital Stereo Signals in High Quality Music Reproduction. University of Erlangen-Nuernberg. 1991.

(Continued)

Primary Examiner—Vivian Chin
Assistant Examiner—Douglas Suthers
(74) *Attorney, Agent, or Firm*—Michael A. Glenn; Glenn Patent Group

(57) **ABSTRACT**

The present invention provides improvements to prior art audio codecs that generate a stereo-illusion through post-processing of a received mono signal. These improvements are accomplished by extraction of stereo-image describing parameters at the encoder side, which are transmitted and subsequently used for control of a stereo generator at the decoder side. Furthermore, the invention bridges the gap between simple pseudo-stereo methods, and current methods of true stereo-coding, by using a new form of parametric stereo coding. A stereo-balance parameter is introduced, which enables more advanced stereo modes, and in addition forms the basis of a new method of stereo-coding of spectral envelopes, of particular use in systems where guided HFR (High Frequency Reconstruction) is employed. As a special case, the application of this stereo-coding scheme in scalable HFR-based codecs is described.

34 Claims, 4 Drawing Sheets
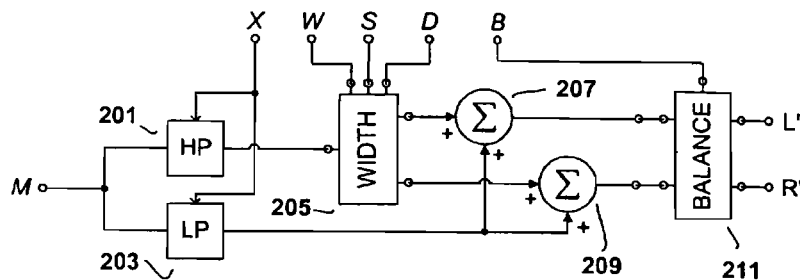
U.S. PATENT DOCUMENTS

5,613,035	A		3/1997	Kim
5,862,228	A	*	1/1999	Davis 381/17
5,873,065	A		2/1999	Akagiri et al.
5,890,108	A		3/1999	Yeldener
2002/0037086	A1	*	3/2002	Irwan et al. 381/307

FOREIGN PATENT DOCUMENTS

EP	0858067	A2	8/1998
EP	1 107 232	A2	6/2001
GB	2100430	A	12/1982
JP	2-177782		7/1990
JP	3-214956		9/1991
JP	9-500252		1/1997
JP	9501286		2/1997
KR	96 03455		3/1996
KR	0103688		3/1996
KR	96 12475		9/1996
KR	0110475		9/1996

OTHER PUBLICATIONS

Herre, Jurgen, et al., "Intensity Stereo Coding," Feb. 26, 1994, Preprints of Papers Presented at the Audio Engineering Society Convention, XP009025131, vol. 96, No. 3799, pp. 1-10.
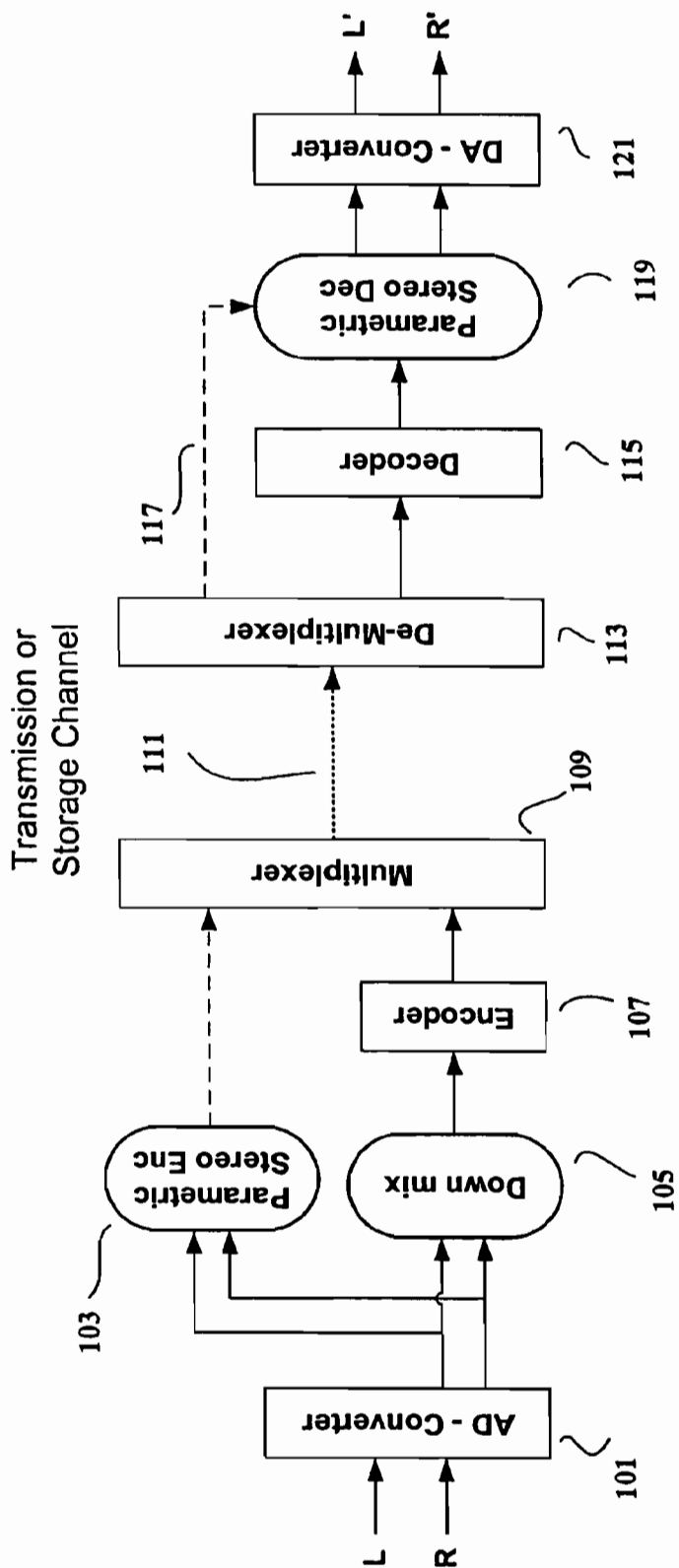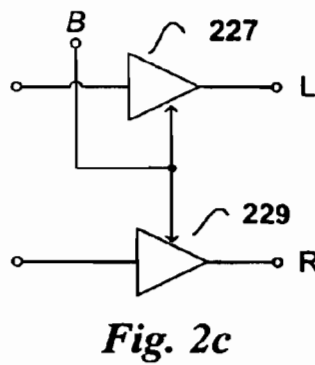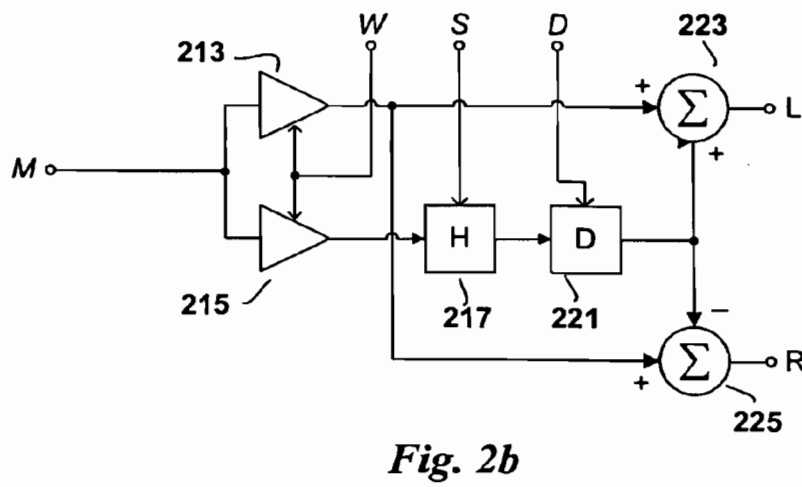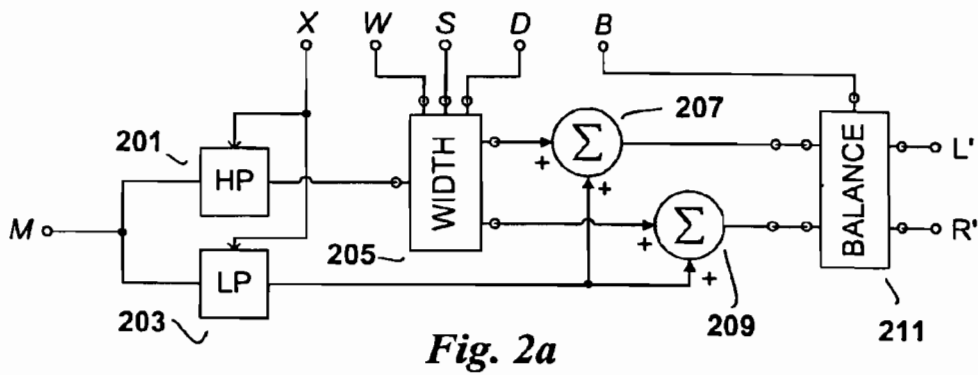
* cited by examiner

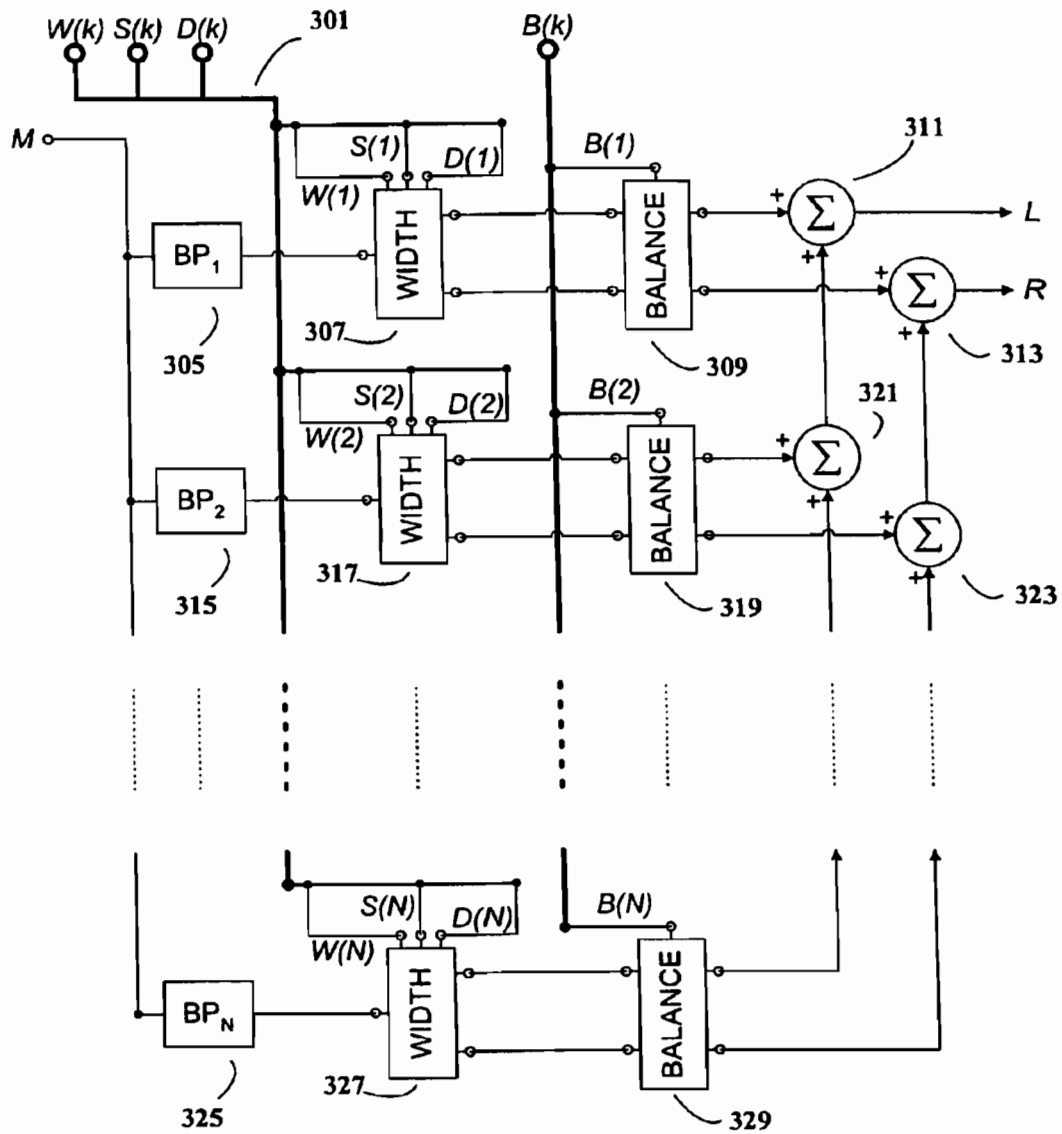*Fig. 1*

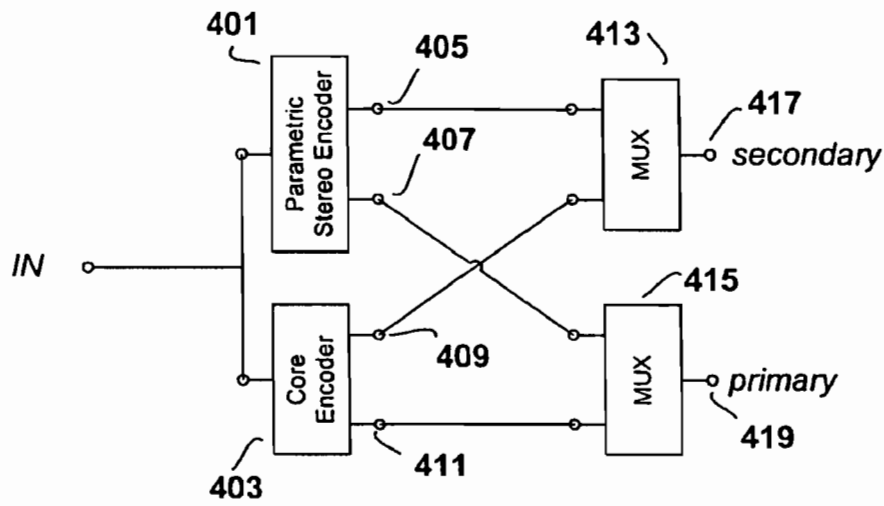*Fig. 2a*

Fig. 2b

Fig. 2c
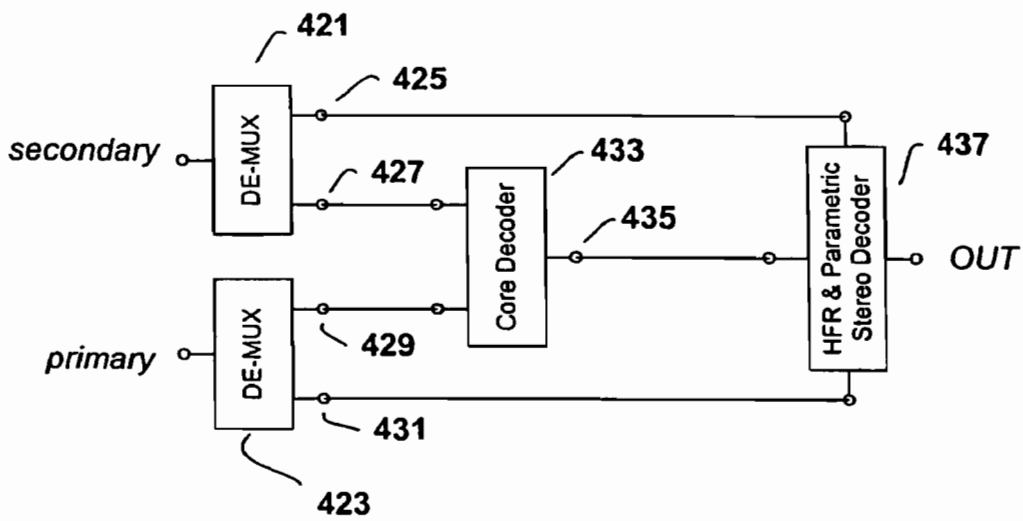
Fig. 3

Fig. 4a

Fig. 4b

EFFICIENT AND SCALABLE PARAMETRIC STEREO CODING FOR LOW BITRATE AUDIO CODING APPLICATIONS

TECHNICAL FIELD

The present invention relates to low bitrate audio source coding systems. Different parametric representations of stereo properties of an input signal are introduced, and the application thereof at the decoder side is explained, ranging from pseudo-stereo to full stereo coding of spectral envelopes, the latter of which is especially suited for HFR based codecs.

BACKGROUND OF THE INVENTION

Audio source coding techniques can be divided into two classes: natural audio coding and speech coding. At medium to high bitrates, natural audio coding is commonly used for speech and music signals, and stereo transmission and reproduction is possible. In applications where only low bitrates are available, e.g. Internet streaming audio targeted at users with slow telephone modem connections, or in the emerging digital AM broadcasting systems, mono coding of the audio program material is unavoidable. However, a stereo impression is still desirable, in particular when listening with headphones, in which case a pure mono signal is perceived as originating from "within the head", which can be an unpleasant experience.

One approach to address this problem is to synthesize a stereo signal at the decoder side from a received pure mono signal. Throughout the years, several different "pseudo-stereo" generators have been proposed. For example in [U.S. Pat. No. 5,883,962], enhancement of mono signals by means of adding delayed/phase shifted versions of a signal to the unprocessed signal, thereby creating a stereo illusion, is described. Hereby the processed signal is added to the original signal for each of the two outputs at equal levels but with opposite signs, ensuring that the enhancement signals cancel if the two channels are added later on in the signal path. In [PCT WO 98/57436] a similar system is shown, albeit without the above mono-compatibility of the enhanced signal. Prior art methods have in common that they are applied as pure post-processes. In other words, no information on the degree of stereo-width, let alone position in the stereo sound stage, is available to the decoder. Thus, the pseudo-stereo signal may or may not have a resemblance of the stereo character of the original signal. A particular situation where prior art systems fall short, is when the original signal is a pure mono signal, which often is the case for speech recordings. This mono signal is blindly converted to a synthetic stereo signal at the decoder, which in the speech case often causes annoying artifacts, and may reduce the clarity and speech intelligibility.

Other prior art systems, aiming at true stereo transmission at low bitrates, typically employ a sum and difference coding scheme. Thus, the original left (L) and right (R) signals are converted to a sum signal, $S=(L+R)/2$, and a difference signal, $D=(L-R)/2$, and subsequently encoded and transmitted. The receiver decodes the S and D signals, whereupon the original L/R-signal is recreated through the operations $L=S+D$, and $R=S-D$. The advantage of this, is that very often a redundancy between L and R is at hand, whereby the information in D to be encoded is less, requiring fewer bits, than in S. Clearly, the extreme case is a pure mono signal, i.e. L and R are identical. A traditional LR-codec encodes this mono signal twice, whereas a S/D codec detects this

redundancy, and the D signal does (ideally) not require any bits at all. Another extreme is represented by the situation where R=−L, corresponding to "out of phase" signals. Now, the S signal is zero, whereas the D signal computes to L. Again, the S/D-scheme has a clear advantage to standard L/R-coding. However, consider the situation where e.g. R=0 during a passage, which was not uncommon in the early days of stereo recordings. Both S and D equal L/2, and the S/D-scheme does not offer any advantage. On the contrary, L/R-coding handles this very well: The R signal does not require any bits. For this reason, prior art codecs employ adaptive switching between those two coding schemes, depending on what method that is most beneficial to use at a given moment. The above examples are merely theoretical (except for the dual mono case, which is common in speech only programs). Thus, real world stereo program material contains significant amounts of stereo information, and even if the above switching is implemented, the resulting bitrate is often still too high for many applications. Furthermore, as can be seen from the resynthesis relations above, very coarse quantization of the D signal in an attempt to further reduce the bitrate is not feasible, since the quantization errors translate to non-neglectable level errors in the L and R signals.

SUMMARY OF THE INVENTION

The present invention employs detection of signal stereo properties prior to coding and transmission. In the simplest form, a detector measures the amount of stereo perspective that is present in the input stereo signal. This amount is then transmitted as a stereo width parameter, together with an encoded mono sum of the original signal. The receiver decodes the mono signal, and applies the proper amount of stereo-width, using a pseudo-stereo generator, which is controlled by said parameter. As a special case, a mono input signal is signaled as zero stereo width, and correspondingly no stereo synthesis is applied in the decoder. According to the invention, useful measures of the stereo-width can be derived e.g. from the difference signal or from the cross-correlation of the original left and right channel. The value of such computations can be mapped to a small number of states, which are transmitted at an appropriate fixed rate in time, or on an as-needed basis. The invention also teaches how to filter the synthesized stereo components, in order to reduce the risk of unmasking coding artifacts which typically are associated with low bitrate coded signals.

Alternatively, the overall stereo-balance or localization in the stereo field is detected in the encoder. This information, optionally together with the above width-parameter, is efficiently transmitted as a balance-parameter, along with the encoded mono signal. Thus, displacements to either side of the sound stage can be recreated at the decoder, by correspondingly altering the gains of the two output channels. According to the invention, this stereo-balance parameter can be derived from the quotient of the left and right signal powers. The transmission of both types of parameters requires very few bits compared to full stereo coding, whereby the total bitrate demand is kept low. In a more elaborate version of the invention, which offers a more accurate parametric stereo depiction, several balance and stereo-width parameters are used, each one representing separate frequency bands.

The balance-parameter generalized to a per frequency-band operation, together with a corresponding per band operation of a level-parameter, calculated as the sum of the left and right signal powers, enables a new, arbitrary

US 7,382,886 B2

3

detailed, representation of the power spectral density of a stereo signal. A particular benefit of this representation, in addition to the benefits from stereo redundancy that also S/D-systems take advantage of, is that the balance-signal can be quantized with less precision than the level ditto, since the quantization error, when converting back to a stereo spectral envelope, causes an "error in space", i.e. perceived localization in the stereo panorama, rather than an error in level. Analogous to a traditional switched L/R- and S/D-system, the level/balance-scheme can be adaptively switched off, in favor of a levelL/levelR-signal, which is more efficient when the overall signal is heavily offset towards either channel. The above spectral envelope coding scheme can be used whenever an efficient coding of power spectral envelopes is required, and can be incorporated as a tool in new stereo source codecs. A particularly interesting application is in HFR systems that are guided by information about the original signal highband envelope. In such a system, the lowband is coded and decoded by means of an arbitrary codec, and the highband is regenerated at the decoder using the decoded lowband signal and the transmitted highband envelope information [PCT WO 98/57436]. Furthermore, the possibility to build a scalable HFR-based stereo codec is offered, by locking the envelope coding to level/balance operation. Hereby the level values are fed into the primary bitstream, which, depending on the implementation, typically decodes to a mono signal. The balance values are fed into the secondary bitstream, which in addition to the primary bitstream is available to receivers close to the transmitter, taking an IBOC (In-Band On-Channel) digital AM-broadcasting system as an example. When the two bitstreams are combined, the decoder produces a stereo output signal. In addition to the level values, the primary bitstream can contain stereo parameters, e.g. a width parameter. Thus, decoding of this bitstream alone already yields a stereo output, which is improved when both bitstreams are available.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of illustrative examples, not limiting the scope or spirit of the invention, with reference to the accompanying drawings, in which:

FIG. 1 illustrates a source coding system containing an encoder enhanced by a parametric stereo encoder module, and a decoder enhanced by a parametric stereo decoder module.

FIG. 2a is a block schematic of a parametric stereo decoder module,

FIG. 2b is a block schematic of a pseudo-stereo generator with control parameter inputs,

FIG. 2c is a block schematic of a balance adjuster with control parameter inputs,

FIG. 3 is a block schematic of a parametric stereo decoder module using multiband pseudo-stereo generation combined with multiband balance adjustment,

FIG. 4a is a block schematic of the encoder side of a scalable HFR-based stereo codec, employing level/balance-coding of the spectral envelope,

FIG. 4b is a block schematic of the corresponding decoder side.

DESCRIPTION OF PREFERRED EMBODIMENTS

The below-described embodiments are merely illustrative for the principles of the present invention. It is understood

4

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. For the sake of clarity, all below examples assume two channel systems, but apparent to others skilled in the art, the methods can be applied to multichannel systems, such as a 5.1 system.

FIG. 1 shows how an arbitrary source coding system comprising of an encoder, 107, and a decoder, 115, where encoder and decoder operate in monaural mode, can be enhanced by parametric stereo coding according to the invention. Let L and R denote the left and right analog input signals, which are fed to an AD-converter, 101. The output from the AD-converter is converted to mono, 105, and the mono signal is encoded, 107. In addition, the stereo signal is routed to a parametric stereo encoder, 103, which calculates one or several stereo parameters to be described below. Those parameters are combined with the encoded mono signal by means of a multiplexer, 109, forming a bitstream, 111. The bitstream is stored or transmitted, and subsequently extracted at the decoder side by means of a demultiplexer, 113. The mono signal is decoded, 115, and converted to a stereo signal by a parametric stereo decoder, 119, which uses the stereo parameter(s), 117, as control signal(s). Finally, the stereo signal is routed to the DA-converter, 121, which feeds the analog outputs, L' and R'. The topology according to FIG. 1 is common to a set of parametric stereo coding methods which will be described in detail, starting with the less complex versions.

One method of parameterization of stereo properties according to the present invention, is to determine the original signal stereo-width at the encoder side. A first approximation of the stereo-width is the difference signal, $D=L-R$, since, roughly put, a high degree of similarity between L and R computes to a small value of D, and vice versa. A special case is dual mono, where $L=R$ and thus $D=0$. Thus, even this simple algorithm is capable of detecting the type of mono input signal commonly associated with news broadcasts, in which case pseudo-stereo is not desired. However, a mono signal that is fed to L and R at different levels does not yield a zero D signal, even though the perceived width is zero. Thus, in practice more elaborate detectors might be required, employing for example cross-correlation methods. One should make sure that the value describing the left-right difference or correlation in some way is normalized with the total signal level, in order to achieve a level independent detector. A problem with the aforementioned detector is the case when mono speech is mixed with a much weaker stereo signal e.g. stereo noise or background music during speech-to-music/music-to-speech transitions. At the speech pauses the detector will then indicate a wide stereo signal. This is solved by normalizing the stereo-width value with a signal containing information of previous total energy level e.g., a peak decay signal of the total energy. Furthermore, to prevent the stereo-width detector from being trigged by high frequency noise or channel different high frequency distortion, the detector signals should be pre-filtered by a low-pass filter, typically with a cutoff frequency somewhere above a voice's second formant, and optionally also by a high-pass filter to avoid unbalanced signal-offsets or hum. Regardless of detector type, the calculated stereo-width is mapped to a finite set of values, covering the entire range, from mono to wide stereo.

FIG. 2a gives an example of the contents of the parametric stereo decoder introduced in FIG. 1. The block denoted

'balance', 211, controlled by parameter B, will be described later, and should be regarded as bypassed for now. The block denoted 'width', 205, takes a mono input signal, and synthetically recreates the impression of stereo width, where the amount of width is controlled by the parameter W. The optional parameters S and D will be described later. According to the invention, a subjectively better sound quality can often be achieved by incorporating a crossover filter comprising of a low-pass filter, 203, and a high-pass filter, 201, in order to keep the low frequency range "tight" and unaffected. Hereby only the output from the high-pass filter is routed to the width block. The stereo output from the width block is added to the mono output from the low-pass filter by means of 207 and 209, forming the stereo output signal.

Any prior art pseudo-stereo generator can be used for the width block, such as those mentioned in the background section, or a Schroeder-type early reflection simulating unit (multitap delay) or reverberator. FIG. 2b gives an example of a pseudo-stereo generator, fed by a mono signal M. The amount of stereo-width is determined by the gain of 215, and this gain is a function of the stereo-width parameter, W. The higher the gain, the wider the stereo-impression, a zero gain corresponds to pure mono reproduction. The output from 215 is delayed, 221, and added, 223 and 225, to the two direct signal instances, using opposite signs. In order not to significantly alter the overall reproduction level when changing the stereo-width, a compensating attenuation of the direct signal can be incorporated, 213. For example, if the gain of the delayed signal is G, the gain of the direct signal can be selected as $\mathrm{sqrt}(1-G^2)$. According to the invention, a high frequency roll-off can be incorporated in the delay signal path, 217, which helps avoiding pseudo-stereo caused unmasking of coding artifacts. Optionally, crossover filter, roll-off filter and delay parameters can be sent in the bitstream, offering more possibilities to mimic the stereo properties of the original signal, as also shown in FIGS. 2a and 2b as the signals X, S and D. If a reverberation unit is used for generating a stereo signal, the reverberation decay might sometimes be unwanted after the very end of a sound. These unwanted reverb-tails can however easily be attenuated or completely removed by just altering the gain of the reverb signal. A detector designed for finding sound endings can be used for that purpose. If the reverberation unit generates artifacts at some specific signals e.g., transients, a detector for those signals can also be used for attenuating the same.

An alternative method of detecting stereo-properties according to the invention, is described as follows. Again, let L and R denote the left and right input signals. The corresponding signal powers are then given by $P_L \sim L^2$ and $P_R \sim R^2$. Now, a measure of the stereo-balance can be calculated as the quotient of the two signal powers, or more specifically as $B = (P_L + e)/(P_R + e)$, where e is an arbitrary, very small number, which eliminates division by zero. The balance parameter, B, can be expressed in dB given by the relation $B_{dB} = 10\log_{10}(B)$. As an example, the three cases $P_L = 10P_R$, $P_L = P_R$, and $P_L = 0.1P_R$ correspond to balance values of +10 dB, 0 dB, and −10 dB respectively. Clearly, those values map to the locations "left", "center", and "right". Experiments have shown that the span of the balance parameter can be limited to for example +/−40 dB, since those extreme values are already perceived as if the sound originates entirely from one of the two loudspeakers or headphone drivers. This limitation reduces the signal space to cover in the transmission, thus offering bitrate reduction. Furthermore, a progressive quantization scheme can be used,

whereby smaller quantization steps are used around zero, and larger steps towards the outer limits, which further reduces the bitrate. Often the balance is constant over time for extended passages. Thus, a last step to significantly reduce the number of average bits needed can be taken: After transmission of an initial balance value, only the differences between consecutive balance values are transmitted, whereby entropy coding is employed. Very commonly, this difference is zero, which thus is signaled by the shortest possible codeword. Clearly, in applications where bit errors are possible, this delta coding must be reset at an appropriate time interval, in order to eliminate uncontrolled error propagation.

The most rudimental decoder usage of the balance parameter, is simply to offset the mono signal towards either of the two reproduction channels, by feeding the mono signal to both outputs and adjusting the gains correspondingly, as illustrated in FIG. 2c, blocks 227 and 229, with the control signal B. This is analogous to turning the "panorama" knob on a mixing desk, synthetically "moving" a mono signal between the two stereo speakers.

The balance parameter can be sent in addition to the above described width parameter, offering the possibility to both position and spread the sound image in the sound-stage in a controlled manner, offering flexibility when mimicking the original stereo impression. One problem with combining pseudo stereo generation, as mentioned in a previous section, and parameter controlled balance, is unwanted signal contribution from the pseudo stereo generator at balance positions far from center position. This is solved by applying a mono favoring function on the stereo-width value, resulting in a greater attenuation of the stereo-width value at balance positions at extreme side position and less or no attenuation at balance positions close to the center position.

The methods described so far, are intended for very low bitrate applications. In applications where higher bitrates are available, it is possible to use more elaborate versions of the above width and balance methods. Stereo-width detection can be made in several frequency bands, resulting in individual stereo-width values for each frequency band. Similarly, balance calculation can operate in a multiband fashion, which is equivalent to applying different filter-curves to two channels that are fed by a mono signal. FIG. 3 shows an example of a parametric stereo decoder using a set of N pseudo-stereo generators according to FIG. 2b, represented by blocks 307, 317 and 327, combined with multiband balance adjustment, represented by blocks 309, 319 and 329, as described in FIG. 2c. The individual passbands are obtained by feeding the mono input signal, M, to a set of bandpass filters, 305, 315 and 325. The bandpass stereo outputs from the balance adjusters are added, 311, 321, 313, 323, forming the stereo output signal, L and R. The formerly scalar width- and balance parameters are now replaced by the arrays W(k) and B(k). In FIG. 3, every pseudo-stereo generator and balance adjuster has unique stereo parameters. However, in order to reduce the total amount of data to be transmitted or stored, parameters from several frequency bands can be averaged in groups at the encoder, and this smaller number of parameters be mapped to the corresponding groups of width and balance blocks at the decoder. Clearly, different grouping schemes and lengths can be used for the arrays W(k) and B(k). S(k) represents the gains of the delay signal paths in the width blocks, and D(k) represents the delay parameters. Again, S(k) and D(k) are optional in the bitstream.

The parametric balance coding method can, especially for lower frequency bands, give a somewhat unstable behavior,

due to lack of frequency resolution, or due to too many sound events occurring in one frequency band at the same time but at different balance positions. Those balance-glitches are usually characterized by a deviant balance value during just a short period of time, typically one or a few consecutive values calculated, dependent on the update rate. In order to avoid disturbing balance-glitches, a stabilization process can be applied on the balance data. This process may use a number of balance values before and after current time position, to calculate the median value of those. The median value can subsequently be used as a limiter value for the current balance value i.e., the current balance value should not be allowed to go beyond the median value. The current value is then limited by the range between the last value and the median value. Optionally, the current balance value can be allowed to pass the limited values by a certain overshoot factor. Furthermore, the overshoot factor, as well as the number of balance values used for calculating the median, should be seen as frequency dependent properties and hence be individual for each frequency band.

At low update ratios of the balance information, the lack of time resolution can cause failure in synchronization between motions of the stereo image and the actual sound events. To improve this behavior in terms of synchronization, an interpolation scheme based on identifying sound events can be used. Interpolation here refers to interpolations between two, in time consecutive balance values. By studying the mono signal at the receiver side, information about beginnings and ends of different sound events can be obtained. One way is to detect a sudden increase or decrease of signal energy in a particular frequency band. The interpolation should after guidance from that energy envelope in time make sure that the changes in balance position should be performed preferably during time segments containing little signal energy. Since human ear is more sensitive to entries than trailing parts of a sound, the interpolation scheme benefits from finding the beginning of a sound by e.g., applying peak-hold to the energy and then let the balance value increments be a function of the peak-holded energy, where a small energy value gives a large increment and vice versa. For time segments containing uniformly distributed energy in time i.e., as for some stationary signals, this interpolation method equals linear interpolation between the two balance values. If the balance values are quotients of left and right energies, logarithmic balance values are preferred, for left-right symmetry reasons. Another advantage of applying the whole interpolation algorithm in the logarithmic domain is the human ear's tendency of relating levels to a logarithmic scale.

Also, for low update ratios of the stereo-width gain values, interpolation can be applied to the same. A simple way is to interpolate linearly between two in time consecutive stereo-width values. More stable behavior of the stereo-width can be achieved by smoothing the stereo-width gain values over a longer time segment containing several stereo-width parameters. By utilizing smoothing with different attack and release time constants, a system well suited for program material containing mixed or interleaved speech and music is achieved. An appropriate design of such smoothing filter is made using a short attack time constant, to get a short rise-time and hence an immediate response to music entries in stereo, and a long release time, to get a long fall-time. To be able to fast switch from a wide stereo mode to mono, which can be desirable for sudden speech entries, there is a possibility to bypass or reset the smoothing filter by signaling this event. Furthermore, attack time constants,

release time constants and other smoothing filter characteristics can also be signaled by an encoder.

For signals containing masked distortion from a psycho-acoustical codec, one common problem with introducing stereo information based on the coded mono signal is an unmasking effect of the distortion. This phenomenon usually referred as "stereo-unmasking" is the result of non-centered sounds that do not fulfill the masking criterion. The problem with stereo-unmasking might be solved or partly solved by, at the decoder side, introducing a detector aimed for such situations. Known technologies for measuring signal to mask ratios can be used to detect potential stereo-unmasking. Once detected, it can be explicitly signaled or the stereo parameters can just simply be decreased.

At the encoder side, one option, as taught by the invention, is to employ a Hilbert transformer to the input signal, i.e. a 90 degree phase shift between the two channels is introduced. When subsequently forming the mono signal by addition of the two signals, a better balance between a center-panned mono signal and "true" stereo signals is achieved, since the Hilbert transformation introduces a 3 dB attenuation for center information. In practice, this improves mono coding of e.g. contemporary pop music, where for instance the lead vocals and the bass guitar commonly is recorded using a single mono source.

The multiband balance-parameter method is not limited to the type of application described in FIG. 1. It can be advantageously used whenever the objective is to efficiently encode the power spectral envelope of a stereo signal. Thus, it can be used as tool in stereo codecs, where in addition to the stereo spectral envelope a corresponding stereo residual is coded. Let the total power P, be defined by $P = P_L + P_R$, where P_L and P_R are signal powers as described above. Note that this definition does not take left to right phase relations into account. (E.g. identical left and right signals but of opposite signs, does not yield a zero total power.) Analogous to B, P can be expressed in dB as $P_{dB} = 10\log_{10}(P/P_{ref})$, where P_{ref} is an arbitrary reference power, and the delta values be entropy coded. As opposed to the balance case, no progressive quantization is employed for P. In order to represent the spectral envelope of a stereo signal, P and B are calculated for a set of frequency bands, typically, but not necessarily, with bandwidths that are related to the critical bands of human hearing. For example those bands may be formed by grouping of channels in a constant bandwidth filterbank, whereby P_L and P_R are calculated as the time and frequency averages of the squares of the subband samples corresponding to respective band and period in time. The sets P_0, P_1, P_2, ..., P_{N-1} and B_0, B_1, B_2, ..., B_{N-1}, where the subscripts denote the frequency band in an N band representation, are delta and Huffman coded, transmitted or stored, and finally decoded into the quantized values that were calculated in the encoder. The last step is to convert P and B back to P_L and P_R. As easily seen form the definitions of P and B, the reverse relations are (when neglecting e in the definition of B) $P_L = BP/(B+1)$, and $P_R = P/(B+1)$.

One particularly interesting application of the above envelope coding method is coding of highband spectral envelopes for HFR-based codecs. In this case no highband residual signal is transmitted. Instead this residual is derived from the lowband. Thus, there is no strict relation between residual and envelope representation, and envelope quantization is more crucial. In order to study the effects of quantization, let Pq and Bq denote the quantized values of P and B respectively. Pq and Bq are then inserted into the above relations, and the sum is formed:

$P_Lq+P_Rq=BqPq/(Bq+1)+Pq/(Bq+1)=Pq(Bq+1)/(Bq+1)=Pq$. The interesting feature here is that Bq is eliminated, and the error in total power is solely determined by the quantization error in P. This implies that even though B is heavily quantized, the perceived level is correct, assuming that sufficient precision in the quantization of P is used. In other words, distortion in B maps to distortion in space, rather than in level. As long as the sound sources are stationary in the space over time, this distortion in the stereo perspective is also stationary, and hard to notice. As already stated, the quantization of the stereo-balance can also be coarser towards the outer extremes, since a given error in dB corresponds to a smaller error in perceived angle when the angle to the centerline is large, due to properties of human hearing.

When quantizing frequency dependent data e.g., multi band stereo-width gain values or multi band balance values, resolution and range of the quantization method can advantageously be selected to match the properties of a perceptual scale. If such scale is made frequency dependent, different quantization methods, or so called quantization classes, can be chosen for the different frequency bands. The encoded parameter values representing the different frequency bands, should then in some cases, even if having identical values, be interpreted in different ways i.e., be decoded into different values.

Analogous to a switched L/R- to S/D-coding scheme, the P and B signals may be adaptively substituted by the P_L and P_R signals, in order to better cope with extreme signals. As taught by [PCT/SE00/00158], delta coding of envelope samples can be switched from delta-in-time to delta-in-frequency, depending on what direction is most efficient in terms of number of bits at a particular moment. The balance parameter can also take advantage of this scheme: Consider for example a source that moves in stereo field over time. Clearly, this corresponds to a successive change of balance values over time, which depending on the speed of the source versus the update rate of the parameters, may correspond to large delta-in-time values, corresponding to large codewords when employing entropy coding. However, assuming that the source has uniform sound radiation versus frequency, the delta-in-frequency values of the balance parameter are zero at every point in time, again corresponding to small codewords. Thus, a lower bitrate is achieved in this case, when using the frequency delta coding direction. Another example is a source that is stationary in the room, but has a non-uniform radiation. Now the delta-in-frequency values are large, and delta-in-time is the preferred choice.

The P/B-coding scheme offers the possibility to build a scalable HFR-codec, see FIG. 4. A scalable codec is characterized in that the bitstream is split into two or more parts, where the reception and decoding of higher order parts is optional. The example assumes two bitstream parts, hereinafter referred to as primary, **419**, and secondary, **417**, but extension to a higher number of parts is clearly possible. The encoder side, FIG. 4a, comprises of an arbitrary stereo lowband encoder, **403**, which operates on the stereo input signal, IN (the trivial steps of AD-respective DA-conversion are not shown in the figure), a parametric stereo encoder, which estimates the highband spectral envelope, and optionally additional stereo parameters, **401**, which also operates on the stereo input signal, and two multiplexers, **415** and **413**, for the primary and secondary bitstreams respectively. In this application, the highband envelope coding is locked to P/B-operation, and the P signal, **407**, is sent to the primary bitstream by means of **415**, whereas the B signal, **405**, is sent to the secondary bitstream, by means of **413**.

For the lowband codec different possibilities exist: It may constantly operate in S/D-mode, and the S and D signals be sent to primary and secondary bitstreams respectively. In this case, a decoding of the primary bitstream results in a full band mono signal. Of course, this mono signal can be enhanced by parametric stereo methods according to the invention, in which case the stereo-parameter(s) also must be located in the primary bitstream. Another possibility is to feed a stereo coded lowband signal to the primary bitstream, optionally together with highband width- and balance-parameters. Now decoding of the primary bitstream results in true stereo for the lowband, and very realistic pseudo-stereo for the highband, since the stereo properties of the lowband are reflected in the high frequency reconstruction. Stated in another way: Even though the available highband envelope representation or spectral coarse structure is in mono, the synthesized highband residual or spectral fine structure is not. In this type of implementation, the secondary bitstream may contain more lowband information, which when combined with that of the primary bitstream, yields a higher quality lowband reproduction. The topology of FIG. 4 illustrates both cases, since the primary and secondary lowband encoder output signals, **411**, and **409**, connected to **415** and **417** respectively, may contain either of the above described signal types.

The bitstreams are transmitted or stored, and either only **419** or both **419** and **417** are fed to the decoder, FIG. 4b. The primary bitstream is demultiplexed by **423**, into the lowband core decoder primary signal, **429** and the P signal, **431**. Similarly, the secondary bitstream is demultiplexed by **421**, into the lowband core decoder secondary signal, **427**, and the B signal, **425**. The lowband signal(s) is(are) routed to the lowband decoder, **433**, which produces an output, **435**, which again, in case of decoding of the primary bitstream only, may be of either type described above (mono or stereo). The signal **435** feeds the HFR-unit, **437**, wherein a synthetic highband is generated, and adjusted according to P, which also is connected to the HFR-unit. The decoded lowband is combined with the highband in the HFR-unit, and the lowband and/or highband is optionally enhanced by a pseudo-stereo generator (also situated in the HFR-unit), before finally being fed to the system outputs, forming the output signal, OUT. When the secondary bitstream, **417**, is present, the HFR-unit also gets the B signal as an input signal, **425**, and **435** is in stereo, whereby the system produces a full stereo output signal, and pseudo-stereo generators if any, are bypassed.

The invention claimed is:

1. A method for coding of stereo properties of a first channel and a second channel of an input signal, the input signal being a two channel signal or a multichannel signal having the first channel and the second channel, comprising:

calculating a stereo width parameter from the first channel and the second channel, wherein the stereo-width parameter represents a degree of similarity between the first channel and the second channel, and wherein the stereo width-parameter is a value from a finite set of values covering a range between a mono situation and a wide stereo situation between the first channel and the second channel;

calculating a balance-parameter, wherein the balance-parameter represents a localization in a stereo field of the first channel and the second channel, and

transmitting or storing the width parameter and the balance parameter so that a decoder, when receiving the width parameter and the balance parameter is adapted to generate a first output channel and a second output

channel of an output signal, the output signal being a two channel output signal or a multichannel output signal having the first output channel and the second output channel, using the stereo width-parameter to control a stereo-width between the first output channel and the second output channel of the output signal and using the balance-parameter to control a localization in the stereo field between the first output channel and the second output channel of the output signal.

2. A method according to claim 1, further comprising: forming a mono signal from the first channel and the second channel of the input signal by combining the first channel and the second channel.

3. A method according to claim 1, further comprising: encoding a mono signal to obtain an encoded mono signal and multiplexing the encoded mono signal, the balance-parameter and the stereo width-parameter to obtain an output bit stream.

4. A method according to claim 1, in which the step of calculating the width-parameter is performed frequency selective such that the width parameter is a vector, and the elements of the vector correspond to separate frequency bands.

5. A method according to claim 1, in which, in the step of calculating the balance parameter, a power for each channel of the input signal is calculated, and the balance-parameter is calculated from a quotient between the powers.

6. A method according to claim 5, in which the powers and the balance-parameter are vectors, in which every element corresponds to a specific frequency band.

7. A method according to claim 6, further comprising the step of calculating an additional level-parameter as a vector sum of the powers to obtain a representation of a spectral envelope of the input signal.

8. A method according to claim 7, further comprising feeding the level-parameter into a primary bit stream of a scalable HFR-based stereo codec, and feeding the balance-parameter into a secondary bit stream of the codec.

9. A method according to claim 1, further comprising feeding a mono signal derived from the first channel and the second channel of the input signal and the width-parameter into a primary bit stream.

10. A method according to claim 1, further comprising a step of quantizing of the balance-parameter, wherein smaller quantization steps around a center position and larger steps towards outer positions are used.

11. A method according to claim 1, further comprising a step of quantizing the width-parameter and the balance-parameter using a quantization method in terms of resolution and range which, for a multiband system, is frequency dependent.

12. A method according to claim 1 further comprising adaptively delta-coding the balance parameter either in time or in frequency.

13. A method according to claim 2, in which the input signal is passed though a Hilbert transformer prior to forming the mono signal.

14. A method for synthesizing a first output channel and a second output channel of an output signal, the output signal being a two channel output signal or a multichannel output signal having the first output channel and the second output channel, using a stereo width-parameter representing a degree of similarity between a first channel and a second channel of an original signal, the original signal being a two channel signal or a multichannel signal having the first channel and the second channel, the stereo width-parameter being a value from a finite set of values covering a range

between a mono situation and a wide stereo situation between the first channel and the second channel of the original signal, and a balance-parameter, representing a localization in a stereo field of the first channel and the second channel of the original signal, and a mono signal derived from the first channel and the second channel of the original signal, comprising:

parametric stereo decoding for generating a synthesized stereo output signal from the mono signal by using the stereo width-parameter to control a stereo-width between the first output channel and the second output channel of the output signal, and using the balance-parameter to control a localization in the stereo field between the first output channel and the second output channel of the output signal.

15. A method according to claim 14, in which the step of parametric stereo decoding comprises the following steps:

generating a first and a second pseudo stereo signal using the mono signal and the width-parameter to control a stereo-width of the pseudo stereo signal, and

balancing the first and the second pseudo stereo signal using the balance parameter to obtain the synthesized output signal.

16. A method in accordance with claim 15, further comprising the step of low-pass filtering the mono signal to obtain a low pass filtered mono signal and adding the low-pass filtered mono signal to the first and second pseudo stereo signal.

17. A method in accordance with claim 15, further comprising the step of high pass filtering the mono signal to obtain a high pass filtered mono signal and subjecting only the high pass filtered mono signal to the step of generating.

18. A method in accordance with claim 15, in which the step of generating includes adding a delayed or phase shifted version of the mono signal to the unprocessed mono signal.

19. A method in accordance with claim 15, in which the step of generating includes the step of weighting the mono signal in accordance with the width factor and the step of delaying the weighted mono signal by a delay and the step of adding the delayed signal to the mono signal using a first sign to obtain the first pseudo stereo signal and the step of adding the delayed signal to the mono signal using a second sign which is opposite to the first sign to obtain the second pseudo stereo signal.

20. A method in accordance with claim 19, further comprising the applying a compensating attenuation to the mono signal which depends on the width parameter so that a total power level of the pseudo stereo signals is equal to the power level of the mono signal.

21. A method in accordance with claim 18, in which the delayed version of the mono signal is progressively attenuated at higher frequencies before being added.

22. A method in accordance with claim 14, further comprising interpolating between two in time consecutive values of said balance-parameters in a way that the momentary value of the corresponding power of said mono signal controls how steep the momentary interpolation should be.

23. A method in accordance with claim 22, in which the interpolating is performed on balance values represented as logarithmic values.

24. A method in accordance with claim 14, in which values of balance-parameters are limited to a range between a previous balance value, and a balance value extracted from other balance values by a median filter or other filter process, where said range is further extendable by moving the borders of said range by a certain factor.

25. A method in accordance with claim 14, in which the width-parameters are processed by a function that gives smaller values for a balance value that, corresponds to a balance position further from the center position.

26. A method in accordance with claim 14, in which the stereo-width of the output signal is retained by means of a pseudo-stereo generator controlled by the width parameter.

27. A method according to claim 1, in which the step of calculating the width parameter includes calculating a difference signal from the first channel and the second channel of the input signal or calculating a cross-correlation between the first channel and the second channel and mapping the difference signal or the cross-correlation to the value of the finite set of values.

28. A method according to claim 1, in which the step of calculating the width parameter includes normalizing the width value using a signal including information of a previous total energy level.

29. A method according to claim 28, in which the signal is a peak decay signal of the total energy.

30. A method according to claim 1, further comprising the steps of low-pass filtering the input signal or high pass filtering the input signal before calculating the width parameter.

31. A method according to claim 30 in which a cutoff frequency used in the step of low pass filtering is above a voice's second formant, or a cut off frequency used in the step of high pass filtering is set such that unbalanced signal-offsets or hum is avoided.

32. An apparatus for coding of stereo properties of a first and a second channel of an input signal, the input signal being a two channel signal or a multichannel signal having the first and the second channel, comprising:

a parametric stereo encoder for calculating a stereo width parameter from the first channel and the second channel, wherein the stereo-width parameter represents a degree of similarity between the first channel and the second channel, and wherein the stereo width-parameter is a value from a finite set of values covering a range between a mono situation and a wide stereo situation between the first channel and the second channel and for calculating a balance-parameter, wherein the balance-parameter represents a localization in a stereo field of the first channel and the second channel, and

wherein the apparatus for coding is adapted for transmitting or storing the width parameter and the balance parameter so that a decoder, when receiving the width

parameter and the balance parameter, is adapted to generate a first output channel and a second output channel of an output signal, the output signal being a two channel output signal or a multichannel output signal having the first output channel and the second output channel, using the stereo width-parameter to control a stereo-width between the first output channel and the second output channel of the output signal and using the balance-parameter to control a localization in the stereo field between the first output channel and the second output channel of the output signal.

33. An apparatus for synthesizing a first output channel and a second output channel of an output signal, the output signal being a two channel output signal or a multichannel output signal having the first output channel and the second output channel, using a stereo width-parameter representing a degree of similarity between a first channel and a second channel of an original signal, the original signal being a two channel signal or a multichannel signal having the first channel and the second channel, the stereo width-parameter being a value from a finite set of values covering a range between a mono situation and a wide stereo situation between the first channel and the second channel of the original signal, and a balance-parameter, representing a localization in a stereo field of the first channel and the second channel of the original signal, and a mono signal derived from the first channel and the second channel of the original signal, comprising:

a parametric stereo decoder for generating a synthesized stereo output signal from the mono signal by using the stereo width-parameter to control a stereo-width between the first output channel and the second output channel of the output signal, and using the balance-parameter to control a localization in the stereo field between the first output channel and the second output channel of the output signal.

34. Apparatus according to claim 33, in which the parametric stereo decoder comprises:

a pseudo stereo generator for generating a first and a second pseudo stereo signal using the mono signal and the width-parameter to control a stereo-width of the pseudo stereo signal, and

a balancing device for balancing the first and the second pseudo stereo signal using the balance parameter to obtain the synthesized output signal.

*   *   *   *   *